AWARD NUMBER: W81XWH-10-1-0151

TITLE: Screening and Selection of New Antagonists of the RING-Mediated Hdm2/Hdmx Interaction

PRINCIPAL INVESTIGATOR: Julio A. Camerero, Ph.D.

CONTRACTING ORGANIZATION: University of Southern California
Los Angeles, CA 90033

REPORT DATE: March 2011

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
1 March 2011	Annual	1 Mar 2010 – 28 Feb 2011

4. TITLE AND SUBTITLE

Screening and Selection of New Antagonists of the RING-Mediated Hdm2/Hdmx Interaction

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-10-1-0151

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Julio A. Camarero, Ph.D.

E-Mail: jcamarer@usc.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Southern California
Los Angeles, CA 90033

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

During the first period of the grant we have been able:

1) Produce large libraries (106 different cyclotides) of cyclotides in E. coli cells using the loop 2 of the cylotide MCoTI-I (Specific aim #1, Task 2a)

2) We have cloned, expressed and characterized a fluorogenic reporter to screen in-cell cyclotide libraries for inhibitors of the RING-mediated interaction between Mdm2/MdmX (specific aim #1, Task 1a & 1b).

15. SUBJECT TERMS
None provided

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U	UU	126	USAMRMC

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

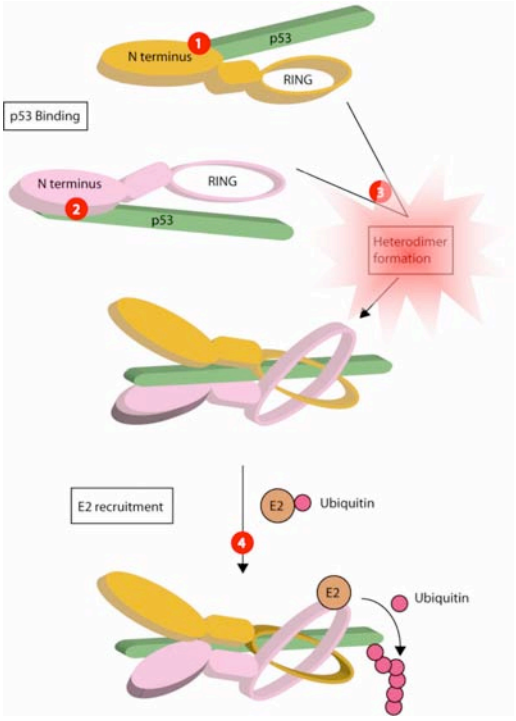
Table of Contents

Introduction

Background

Prostate cancer poses a major public health problem in the United States and worldwide. It has the highest incidence and is the second most common cause of cancer deaths in North American men resulting in over 30,000 deaths per annum. Consequently, there is an urgent need to develop novel therapeutic approaches. The molecular mechanisms of development and progression of prostate cancer are complicated and likely to involve multiple factors. The human double minute 2 (Hdm2) protein is amplified or overexpressed in a number of human tumors, including prostate cancer. Importantly, the Hdm2 antagonist nutlin-3, which is particularly effective in causing p53-dependent apoptosis in Hdm2-amplified cultured cells, exhibits antitumor activity on human prostate LNCaP and other xeno-grafts in nude mice. Hdm2 promotes p53 degradation through an ubiquitin-dependent pathway (Fig. 1). The exact mechanism by which p53 is stabilized is unclear, although a series of post-translational modifications to itself, Hdm2 and the closely related protein Hdmx (also known as Hdm4), are thought to dissociate the p53-Hdm2 complex leading to increased levels of p53. Although Hdmx contains a RING domain that is very similar to the RING domain of Hdm2, it does not possess intrinsic E3 ubiquitin ligase activity. Dimerization, mediated by the conserved C-terminal RING domains of both Hdm2 and Hdmx, appears to greatly augment this activity. While the Hdm RING domains can form homodimers, heterodimers form preferentially resulting in reduced auto-ubiquitylation of Hdm2 and increased p53 ubiquitylation. Thus disruption of this interaction should inactivate Hdm2 E3 ligase activity and consequently increase p53 abundance. The recent elucidation of the structure of the complex formed by the RING domains of Hdm2 and Hdmx suggests the feasibility of obtaining Hdm-specific E3 ligase inhibitors by targeting the Hdm2/HdmX RING domain dimer interface rather than the primary E2 binding site that is common to many RING domain E3-ubiquitin ligases (Fig. 2).

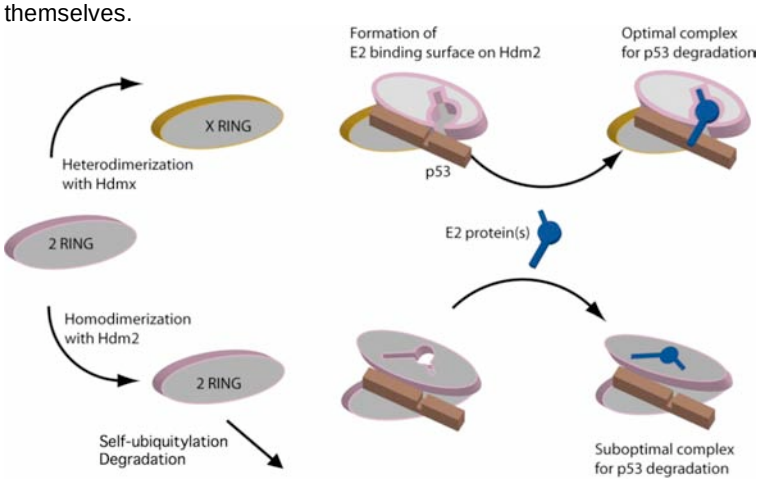
Figure 1. Hdm2-Hdmx may be functionally inhibited at multiple steps to reactivate p53 function. Numbered circles indicate potential therapeutical targets for the development of Hdm2/Hdmx antagonists. In this proposal will target step 3 highlighted in green.



Objectives

Disruption of Hdm2 function is a very novel attractive therapeutic target for prostate cancer [2, 3] (Fig. 1). As mentioned earlier, nutlin-3, a peptidomimetic that disrupts the p53-Hdm2 interaction, activates p53 pathways both in vitro and in vivo in human cell lines that possess wild-type p53 and overexpress Hdm2 [4]. Vousden and colleagues have also established that it is possible to stabilize p53 by directly inhibiting the E3-ligase activity of Hdm2 [6], although this approach may also inhibit other E3s. The recent elucidation of the structure of the Hdm2-Hdmx RING domain heterodimer shows that both protein domains contribute residues for E3-ligase activity (Fig. 3). This strongly suggests that it might be possible to obtain Hdm-specific E3 ligase

Figure 2. Hdm2-Hdmx heterodimers are more effective p53 ubiquitin ligases that Hdm2 homodimers, which can also self-ubiquitylate themselves.



inhibitors by targeting the Hdm2-HdmX RING domain dimer interface rather than the primary E2 binding site that is common to many RING domain E3-ubiquitin ligases.

To achieve this objective we are using cell-based libraries of cyclotides for selecting specific cyclotide sequences able to antagonize the RING-mediated interaction between Hdm2 and Hdmx to obtain Hdm-specific E3 ligase inhibitors. The use of the modified protein splicing technology developed in the Camarero lab allows to generate large, genetically-encoded cyclotide libraries in bacterial cells. These cell-based libraries are the screened using an in-cell FRET-based reporter in combination with high throughput flow cytometry to identify bacteria encoding cyclotides able to disrupt Hdm2-Hdmx interactions. Selected cyclotides are then characterized by NMR and assayed in mammalian cells using the BiLC assay implemented in p53 wild-type and p53-null cancer cell lines to ascertain p53-dependent biological activity.

This proposal represents a novel approach for antagonizing Hdm2-Hdmx E3-ligase activity. Selected cyclotides will be highly specific for antagonizing Hdm2 E3-ligase, and for eliciting p53-dependent cytotoxicity in cancer cells. The use of the cyclotide scaffold will enable these peptide antagonists to have the required increased stability, cellular membrane penetration, proteolysis resistance and serum clearance needed to be considered as viable drug development candidates. It is also important to remark, that this cell-based technology could be easily adapted to screen for antagonists for other relevant protein-protein interactions in prostate cancer; for example those that may induce tumor cell apoptosis independent of p53 or compounds able to reactivate mutant p53. These compounds could be used in combination with Hdm2 antagonists to prevent tumor relapse or secondary tumor formation.

Body

A. Specific Aims

We are using a cyclotide-based molecular scaffold for generating molecular libraries that are screened and selected in vivo for potential antagonists for the RING-mediated Hdm2/Hdmx interaction. In this innovative approach, we are using cell-based libraries (*E. coli* cell libraries) where every single cell will express a different cyclotide, in what we could call a single cell-single compound approach. These compounds are then screened and selected for their ability to inhibit the Hdm2/Hdmx interaction inside the bacterial cell using a genetically-encoded FRET-based reporter [7] in combination with high throughput flow cytometry to identify bacteria encoding cyclotides able to disrupt Hdm2-Hdmx interactions. This screening assay is optimized to be used in *E. coli* in combination fluorescence activated cell sorting (FACS) and is designed to minimize the number of false positives. The Camarero lab has developed a similar assay to screen cyclotides able to inhibit *B. anthracis* Lethal Factor protease activity anthrax toxin binding, demonstrating the feasibility of this approach for use in bacteria.

Selected cyclotides are also characterized by NMR and assayed in mammalian cells using the BiLC assay implemented in p53 wild-type and p53-null cancer cell lines to ascertain p53-dependent biological activity. The BiLC assay developed in the Wahl lab shows a high dynamic range, high degree of reproducibility and will be used to validate the ability of any cyclotide selected in bacteria to enter mammalian cells in concentration sufficient to antagonize Hdm2-Hdmx RING-mediated interaction.

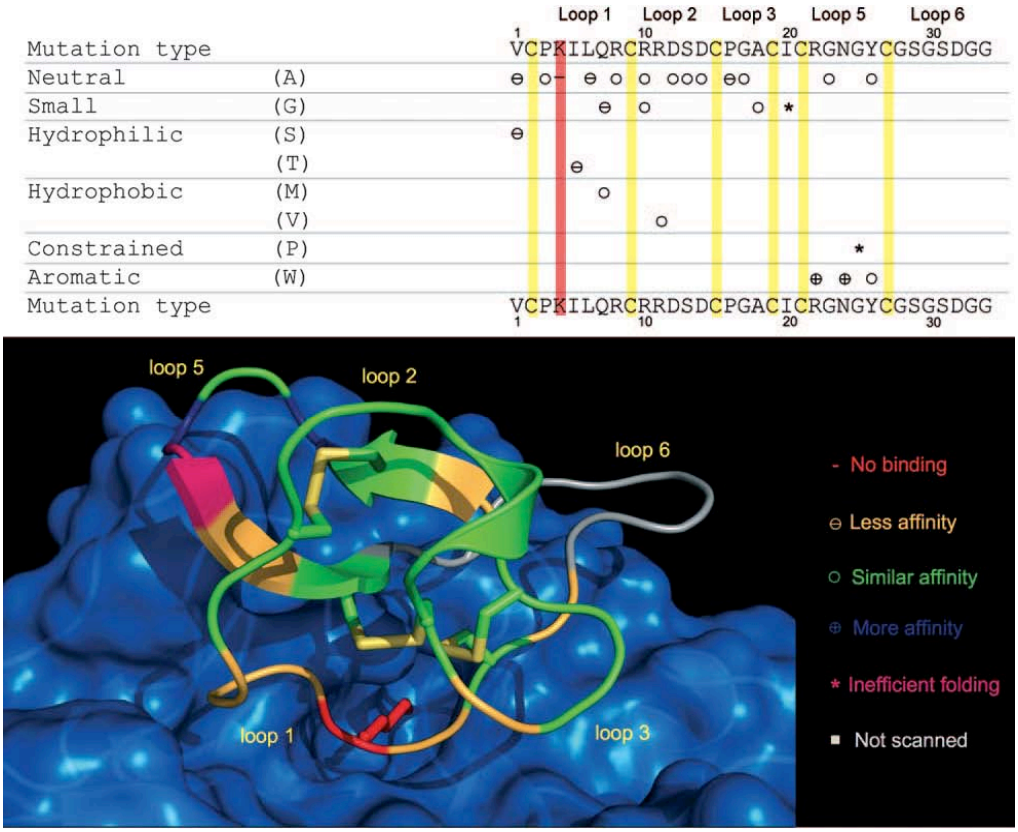
Specific Aim 1. To screen and select cyclotide-based peptides able to disrupt the Hdm2-Hdmx RING heterodimer. The objectives of this aim are the production of large genetically-encoded libraries of cyclotides in living *E. coli* cells (≈10^9) and the development of FRET-based in vivo screening reporter to select cyclotides able to inhibit Hdm2-Hdmx RING heterodimer. Cells able to express active cyclotides will be selected using high throughput flow cytometry methods such as fluorescence activated cell sorting (FACS)

Specific Aim 2. To test and evaluate the inhibitory and biological activity of selected cyclotides. Selected cyclotides will be tested in vitro using a combination of fluorescence assays and nuclear magnetic resonance (NMR). Biological activity will be assayed using different cancer cell lines to evaluate their ability to activate endogenous p53.

B. Studies and Results

1) Biosynthesis and characterization of genetically-encoded cyclotide-based libraries. Our group has recently developed and successfully used a bio-mimetic approach for the biosynthesis of folded cyclotides inside cells by making use of modified protein splicing unit [8]. Using this approach, we have biosynthesized a small

Figure 3. Summary of the relative affinities for trypsin of the different MCoTI-I mutants studied in this work [2]. A model of cyclotide MCoTI-I bound to trypsin is shown at the bottom and indicates the positions of the mutations. The Lys4 side chain is shown in red bound to the specificity pocket of trypsin. The model was produced by homology modeling at the Swiss model workspace (http://swissmodel.expasy.org//SWISS-MODEL.html) by usingthe structure of CPTI-II-trypsin complex (PDB ID: 2btc) as the template.



genetically-encoded library based on the cyclotide MCoTI-I [9] (See paper #7 in the Appendix Section for further details). This cyclotide MCoTI-I is a powerful trypsin inhibitor recently isolated from the seeds of *Momordica cochinchinensis*, a plant member of cucurbitaceae family. In order to explore the contribution of individual residues to biological activity and structural integrity a small library containing multiple amino acid mutants was generated in *E. coli* cells and its activity assayed using a trypsin-binding assay. Using competition-binding experiments we were able to estimate the relative binding propensities of the different mutants. These data provide significant insights into the structural constraints of the MCoTI cyclotide framework and the functional elements for trypsin binding. To our knowledge, this is the first time that the biosynthesis of a genetically-encoded library of MCoTI-based cyclotides containing a comprehensive suite of amino acid mutants is reported. The mutagenesis results obtained in our work highlighting the extreme robustness of the cyclotide scaffold to mutations (Fig. 3). Only two of the 27 mutations studied in the cyclotide MCoTI-I, affected negatively the adoption a native cyclotide fold. Intriguingly, the rest of the mutations allowed the adoption of a native fold as indicated by ES-MS analysis and their ability to bind trypsin. These results should provide an excellent starting point for the effective design of MCoTI-based cyclotide libraries for rapid screening and selection of de novo cyclotide sequences with specific biological activities.

We have also used our unique ability to express folded cyclotides using bacterial expression systems to incorporate NMR-active nuclei such as ^{15}N into folded cyclotides to study the backbone dynamics of the this extraordinary molecular scaffold [10] (See paper # 4 in the Appendix Section for further details). Determination of the backbone dynamics of these fascinating micro-proteins is key for understanding their physical and biological properties. Internal motions of a protein on different time scales, extending from picoseconds to second, have been suggested to play an important role in its biological function. A better understanding of the backbone dynamics of the cyclotide scaffold is extremely helpful for evaluating its utility as a scaffold for developing specific protein-capture reagents. Such insight will help us in the design of optimal focused libraries than can be used for the discovery of new cyclotides sequences with novel biological activities. We have recently reported the backbone dynamics of the cyclotide MCoTI-I in the free state and complexed to its binding partner trypsin in solution [10] (Fig. 4). This is the first time the backbone dynamics of a natively folded cyclotide has been reported in the literature. Our results on the backbone dynamics of free cyclotide MCoTI-I confirm that MCoTI-I adopts a well-folded and highly compact structure with an $\langle S^2 \rangle$ value of 0.83. This value is similar to those found in the regions of well-folded proteins with restricted backbone dynamics. More surprisingly, however, was the fact that the backbone of MCoTI-I, and specially the binding loop, increased ps-ns mobility when bound to trypsin. This increment in backbone mobility may help to minimize the entropic penalties required for binding. This dynamic decoupling between the side-chain terminus from the rest of the

aliphatic part of the side-chain may be a general biophysical strategy for maximizing residual side-chain and potentially backbone conformational entropy in proteins and their complexes. Using these data we have recently produced a combinatorial library using one of the loops (loop 2) of the cyclotide MCoTI-I as molecular scaffold in *E. coli* cells. For this purpose we have used two different expression vectors: pTXB1 and pBAD24. The Mxe Gyrase intein was subcloned into pBAD24 following the procedure depicted in Figure 5. Genetically-encoded cyclotide-based libraries were generated at the DNA level using double stranded (ds) DNA inserts with degenerate sequences for the loop 2 of cyclotide MCoTI-I in combination with PCR. Briefly, a long degenerate synthetic oligonucleotide (which encodes the whole cyclotide, ≈ 100 nt long and the degenerate loop) template is PCR amplified using 5'- and 3'-primers corresponding to the non degenerate flanking regions (Fig. 6A). The degenerate synthetic oligonucleotide template will be synthesized using a NN(G/T) codon scheme for the randomized loops. This scheme uses 32 codons to encode all 20 amino acids and encodes only 1 stop codon. The theoretical for such library is 205 or 3.2 x 10^6 sequences The resulting double stranded degenerate DNA was double digested and then ligated to a linearized intein-containing expression vector to produce a library of pTXB1- or pBAD24-based plasmids (Fig. 6B). These libraries were transformed into electrocompetent *E. coli* cells to finally obtain a library of cells (≈10^6). Characterization of the libraries was carried by picking and sequencing of 50 different clones. The characterization of the library, where loop 2 was completely randomized, revealed a

Figure 4. NMR analysis of the backbone dynamic of free and trypsin bound MCoTI-I. a) {^{15}N, ^{1}H}NMR heteronuclear single quantum correlation (HSQC) spectrum of free MCoTI-I. Chemical shift assignments of the backbone amides are indicated. b) Overlay of the {^{15}N,^{1}H} HSQC spectra of free (black) and trypsin bound MCoTI-I (red). Residues with large average amide chemical shift differences between two different states (>0.3 ppm) are indicated. Peaks that are broadened in trypsin bound MCoTI-I are indicated by grey circles. c) Average amide chemical shift difference for all the assigned residues in free and trypsin bound MCoTI-I.

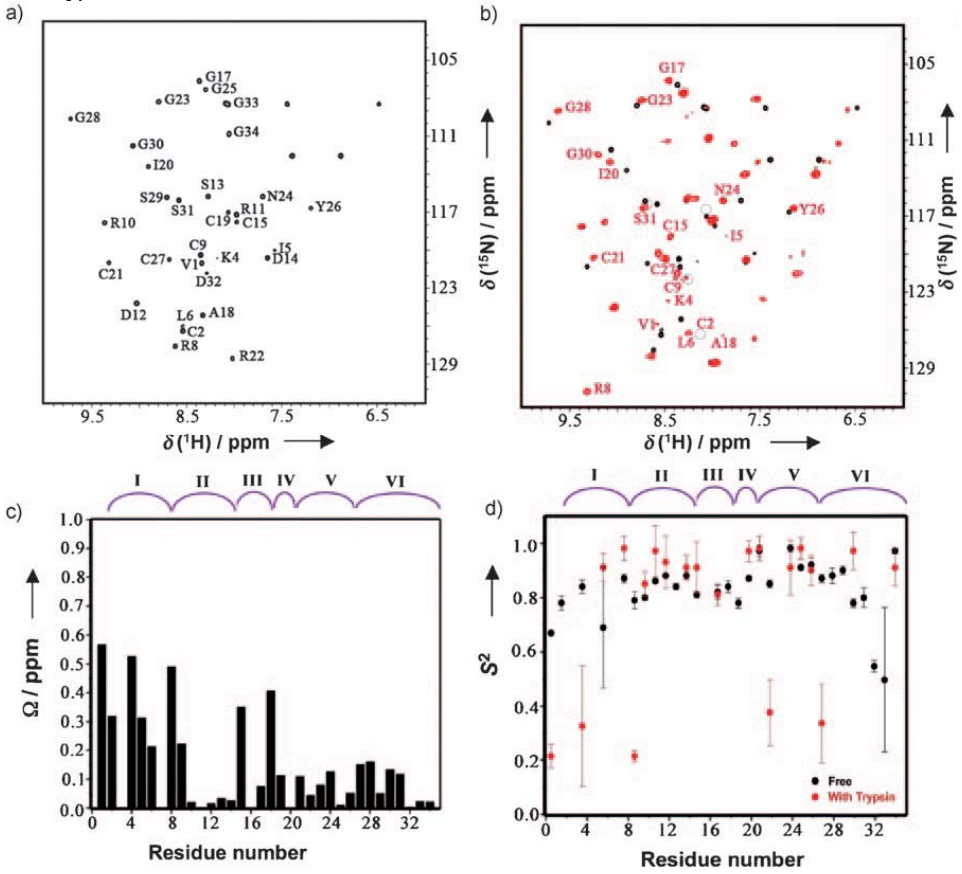
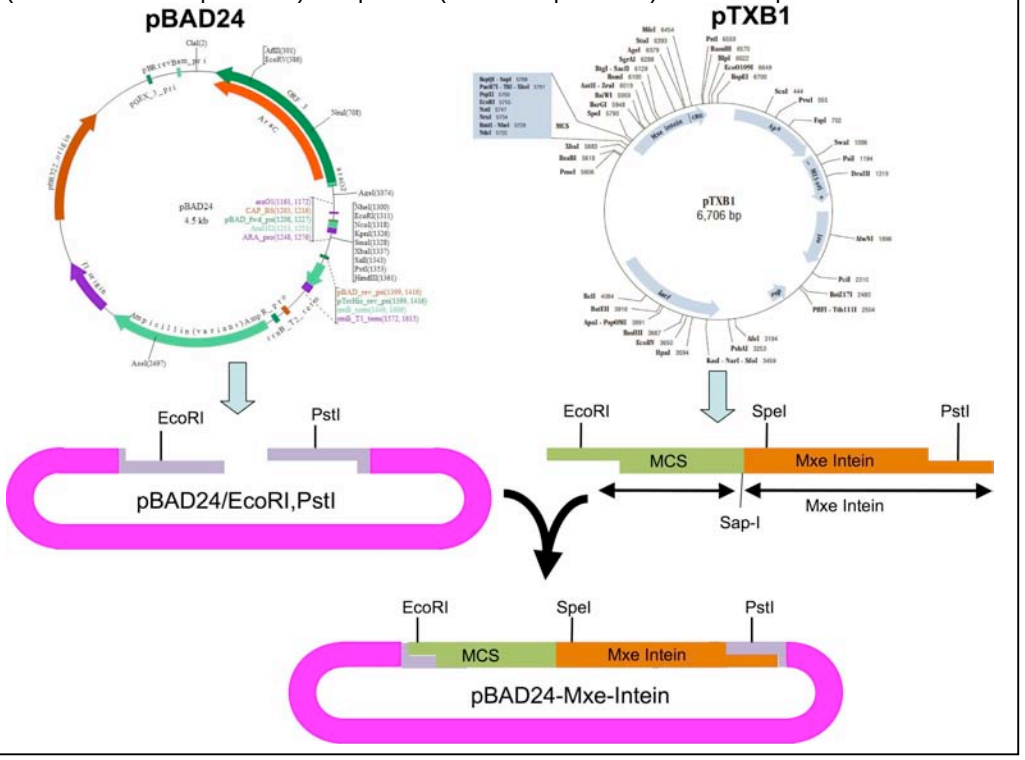


Figure 5. Scheme used for the cloning of cyclotide-based libraries using the pBAD24 (Arabinose driven promoter) and pTXB1 (T7-driven promoter) derived expression vectors.



complexity of ≈0.9x10^6 different unique cyclotide sequences, very close to the theoretical maximum value (3.4x10^6) (Fig. 7). This library has been cloned into two different expression plasmids with different promoters, origin of replication and antibiotic resistance to allow the screening of the libraries using cell-based reporters encoded in orthogonal plasmids. These include T7- and arabinose-driven *E. coli*

Figure 6. Molecular approach to build MCoTI-I-based genetically encoded libraries using loop 2 (A) into orthogonal plasmids pTXB1 and pBAD24-Intein (B).

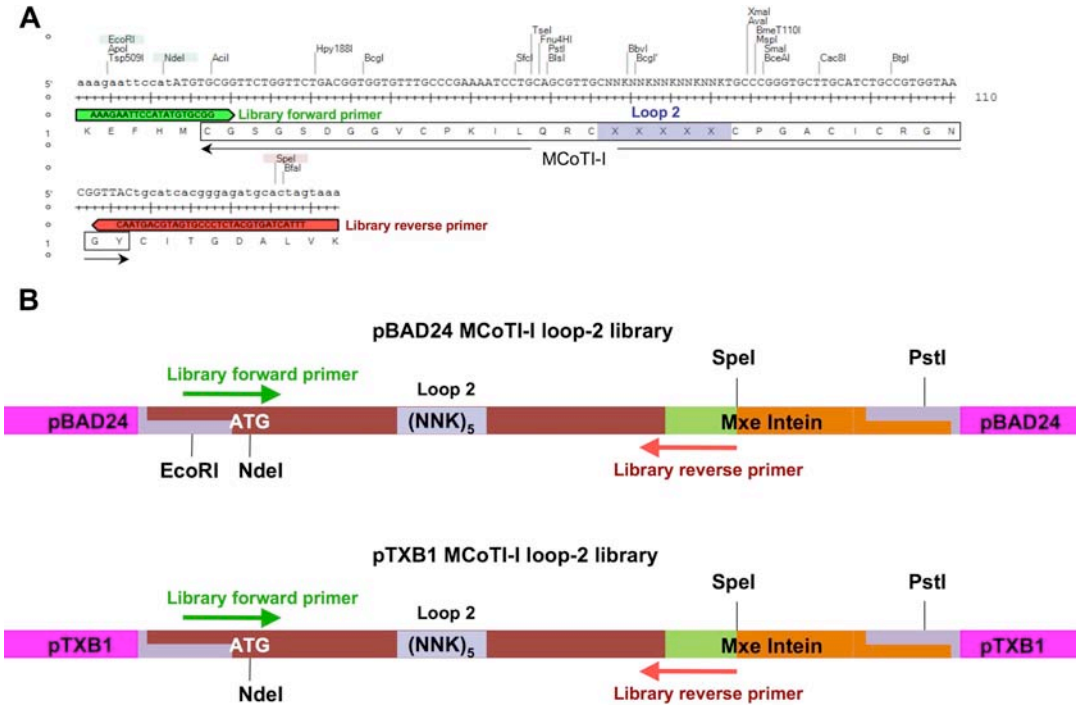
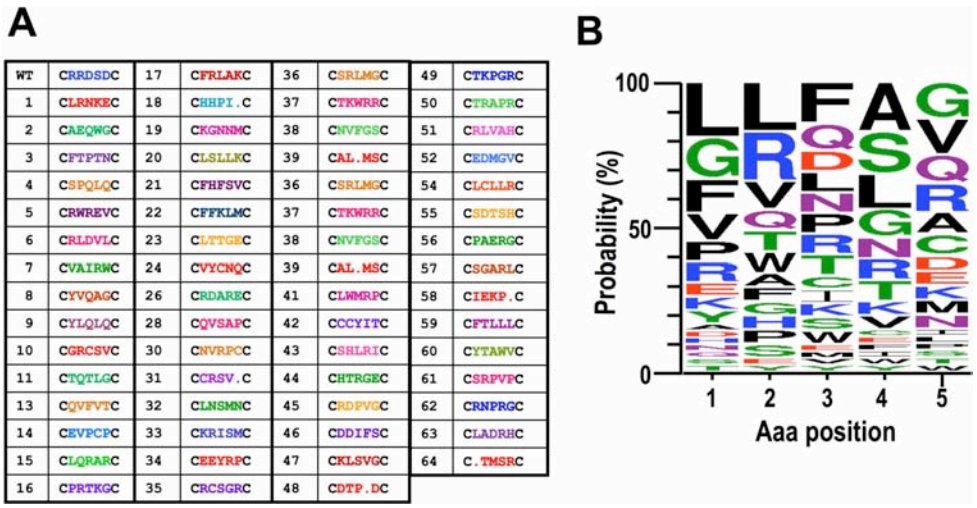


Figure 7. Genomic characterization of the MCoTI-library where loop 2 was completely randomized using an NNK scheme. A. DNA sequences of loop 2 for 60 different clones randomly chosen from the from a MCoTI-I loop 2 based library using the pTXB1 plasmid. B. Graphical representation of the amino acid composition of the 60 clones sequenced. The graph was produced using WebLogo 3 (http://weblogo.threeplusone.com/).

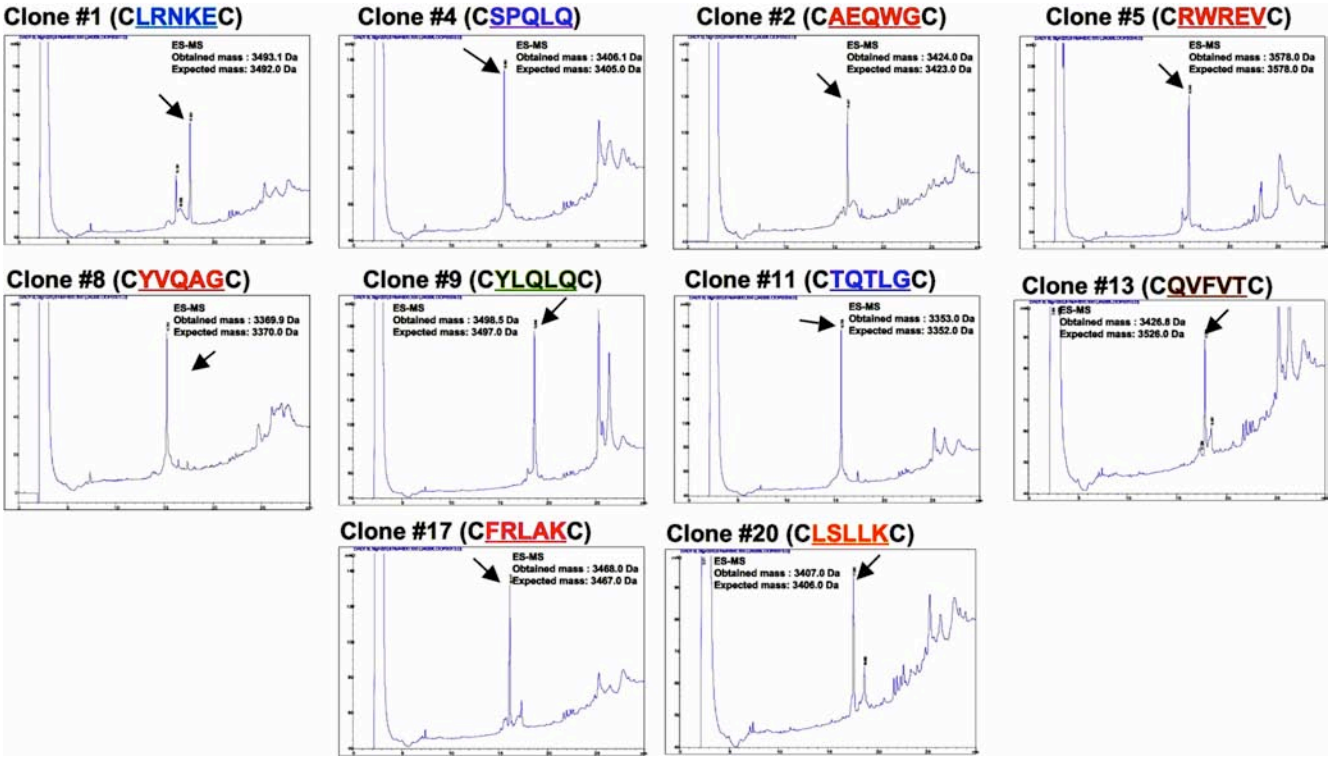


A

WT	CRRDSDC	17	CFRLAKC	36	CSRLMGC	49	CTKPGRC
1	CLRNKEC	18	CHHPI.C	37	CTKWRRC	50	CTRAPRC
2	CAEQWGC	19	CKGNNMC	38	CNVFGSC	51	CRLVAHC
3	CFTPTNC	20	CLSLLKC	39	CAL.MSC	52	CEDMGVC
4	CSPQLQC	21	CFHFSVC	36	CSRLMGC	54	CLCLLRC
5	CRWREVC	22	CFFKLMC	37	CTKWRRC	55	CSDTSHC
6	CRLDVLC	23	CLTTGEC	38	CNVFGSC	56	CPAERGC
7	CVAIRWC	24	CVYCNQC	39	CAL.MSC	57	CSGARLC
8	CYVQAGC	26	CRDAREC	41	CLWMRPC	58	CIEKP.C
9	CYLQLQC	28	CQVSAPC	42	CCCYITC	59	CFTLLLC
10	CGRCSVC	30	CNVRPCC	43	CSHLRIC	60	CYTAWVC
11	CTQTLGC	31	CCRSV.C	44	CHTRGEC	61	CSRPVPC
13	CQVFVTC	32	CLNSMNC	45	CRDPVGC	62	CRNPRGC
14	CEVPCPC	33	CKRISMC	46	CDDIFSC	63	CLADRHC
15	CLQRARC	34	CEEYRPC	47	CKLSVGC	64	C.TMSRC
16	CPRTKGC	35	CRCSGRC	48	CDTP.DC		

expression plasmids (Fig. 5), and are compatible with the plasmids that will be used to express the cell-reporters (see below). We estimated that ≈80% of the loop 2-randomized cyclotides were able to adopt a native cyclotide fold (Fig. 8).

Figure 8. HPLC analytical traces of the cyclization/folding crudes for individual clones isolated from the MCoTI-I loop 2 library expressed using plasmid pTXB1 in *E. coli* before and after purification using trypsin-agarose beads. Folded cyclotides were correctly characterized by ES-MS (mass spectrometry) and their ability to bind trypsin.



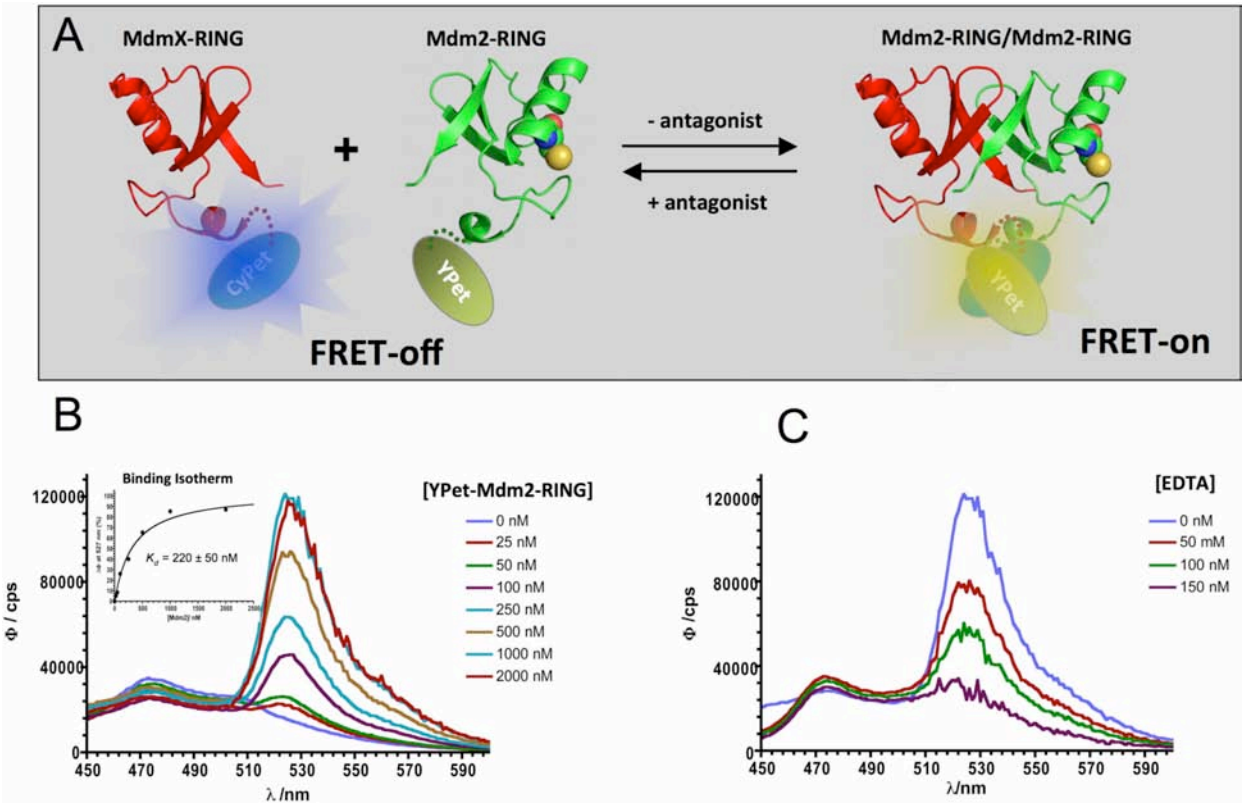
We have also shown that the use of modified protein splicing units can be also used for the generation of cell-based libraries using other cyclic molecular scaffolds thus opening the possibility of using alternative templates for the production of therapeutics able to inhibit molecular interactions relevant to prostate cancer. These include the Bowman-Birk inhibitor SFTI-1 [9] and backbone cyclized α-defensins (manuscript in preparation).

2) Cell-based reporter to screen in-cell RING-mediated Hdm2/HdmX interactions. We have generated a fluorescent reporter to screen RING-Mdm2/RING-MdmX interaction. The principle for this approach is depicted in Figure 9. Our FRET-based reporter system uses a CyPet and YPet fluorescent proteins, which are fused the N-terminus of the RING domains of Mdm2 (429-491 aa) and MdmX (427-490 aa), respectively. The N-terminal region of the Mdm2/X RING domains can easily tolerate the addition of different protein domains or protein fragments without altering their heterodimerization and biological function. For example the Wahl group has shown that half-luciferase fragments can be added without affecting the ability of Hdm2 and Hdmx RING domains to heterodimerize. Moreover, to prevent any potential steric hindrance that could interfere with the molecular recognition process, we initially introduced the flexible linker $[Gly\text{-}Gly\text{-}Ser]_5$ between the interacting proteins or protein domains and the corresponding fluorescent proteins

Both fluorescent protein constructs were successfully cloned and expressed in *E. coli*. As expected, CyPet-Mdm2 and Ypet-MdmX were biologically active in an in vitro fluorescence-binding assay using fluorescence resonance emission transfer (FRET) (Fig. 9A and 9B). When CyPet-MdmX was titrated with increasing amounts of YPet-Mdm2 the FRET signal increased until saturation of the complex was achieved (Fig. 9B). Using these data we were able to plot a binding isotherm for the RING-mediated Mdm2/MdmX interaction and calculate the dissociation constant of this complex (K_d = 220 ± 50 nM, Fig. 9B inset). We are also planning to titrate CyPet-Mdm2 with YPet-MdmX to confirm the value of the affinity constant between the two interacting RING domains (work in progress). As shown in Figure 9C, we were also able to follow the inhibition of the interaction by fluorescence. This was accomplished by titrating a solution containing the CyPet-MdmX/YPet-

Mdm2 complex with ethylenediaminetetraacetic acid (EDTA), a well know inhibitor of protein interactions mediated by proteins containing transition metals. RING domains require 2 atoms of Zn for correctly folding. Our data clearly indicates that the RING-based YPet/CyPet FRET reporter can be efficiently used *in vitro* to screen antagonists against this heteromolecular complex.

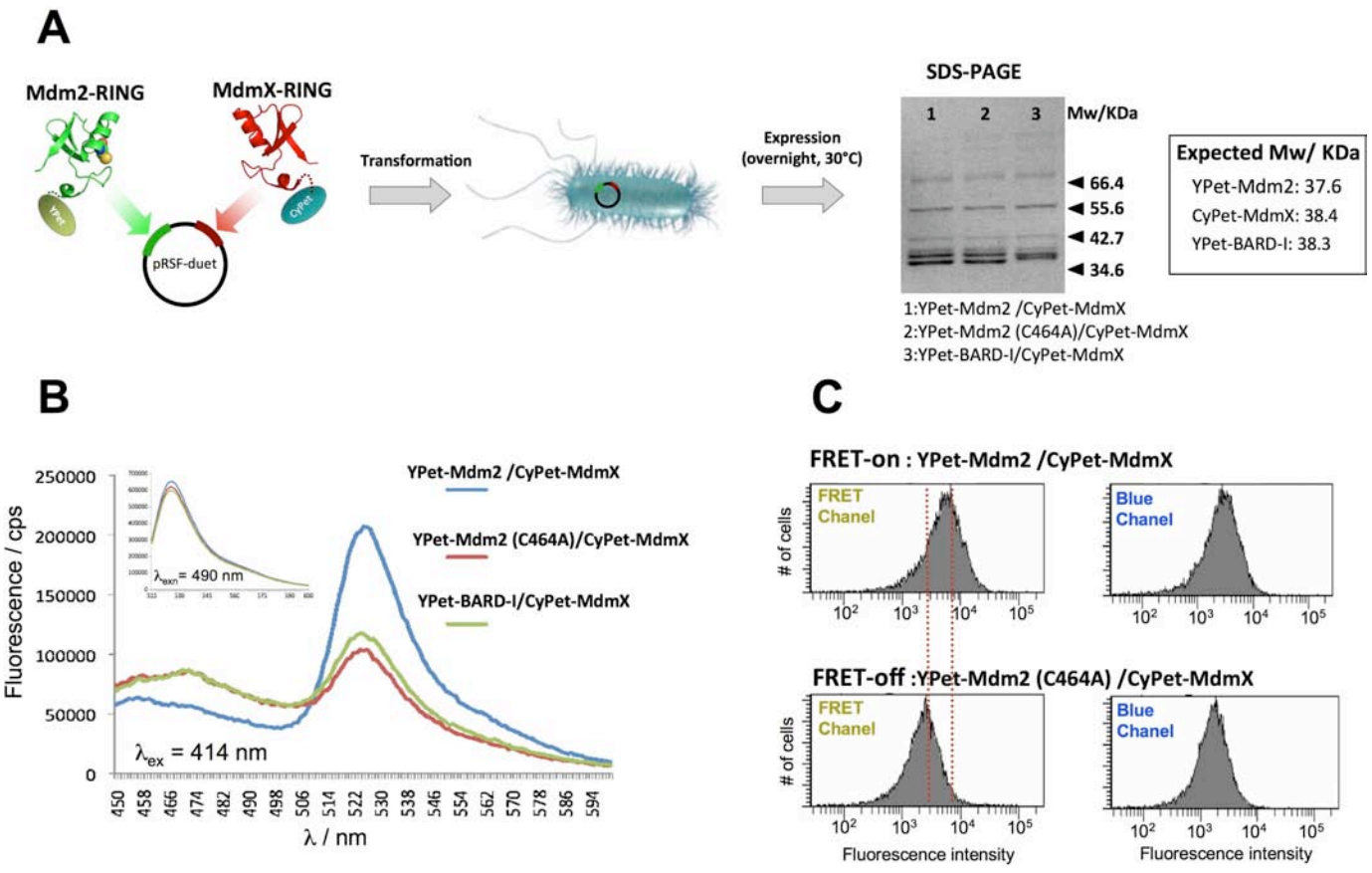
Figure 9. FRET-based reporter for screening Mdm2/MdmX RING-mediated Interactions. A. Principle for the FRET-based reporter to screen for antagonists against the RING-mediated Mdm2/MdmX heteronuclear complex. The formation of the complex brings in close proximity fluorescent proteins YPet and CyPet. Excitation of CyPet with blue light allows the transfer of energy to the yellow fluorescent protein YPet producing yellow fluorescence. Inhibition of the Mdm2/MdmX interaction prevents the transfer of energy from the blue fluorescent protein CyPet thus giving only blue fluorescence. B. Titration of CyPet-MdmX with increasing amounts of YPet-Mdm2. The formation of the heteronuclear complex can be followed by the increase in yellow fluorescence. Plotting the increase increase in yellow fluorescence versus the concentration of YPet-Mmd2 (inset) provides a binding isotherm that allows the determination of the affinity constant between the RING domains of Mdm2 and MdmX. C. The destruction of the Mdm2/MdmX complex by adding increasing amounts of EDTA can be also flowed by the decrease in yellow fluorescence.



Next we tested the ability to use it inside living *E. coli* cells. For that purpose the FRET-pairs CyPet-MdmX/YPet-Mdm2 or CyPet-Mdm2/YPet-MdmX were cloned into a pRSF duet vector and expressed in E. coli cells (Fig. 10). The pRSF duet allows poly-cistronic expression of the corresponding YPet-CyPEt FRET pair, and is orthogonal to the pBAD and pTXB1 vectors (both used for the construction of the cyclotide-based libraries). As shown in Figure 10, when the Hdm2-Hdmx RING heterodimer is formed in vivo the pair CyPEt-YPet exhibits high FRET signal indicating the formation of the complex. Background FRET signal was evaluated using cells expressing YPet-MdmX and the CyPet-Mdm2 mutant (C464A). This mutation in the RING domain of Mdm2 prevents Hdm2-Hdmx heterodimerization and provides the background FRET signal when the Mdm2/MdmX complex is not formed. We also tested the FRET background using a cell line co-expressing the RING domain of BARD-I (25-189) fused to YPet at its N-terminus with CyPET-MdmX as these two RING domains do not interact with each other. In both cases the background FRET signal when the RING heterodimerization was prevented (i.e. FRET-off state) was 200% smaller that of the positive control using CyPet-MdmX/YPet-Mdm2 or CyPet-Mdm2/YPet-MdmX (i.e. FRET-on state) (Fig. 10B). Analysis by

fluorescence activated cell sorting (FACS) also revealed that both populations of cells, i.e. FRET-on and FRET-off cells can be easily separated by FACS (Fig. 10C). We anticipate that this in vivo FRET-reporter will allow the easy separation of positive clones by FACS. AT this time we are also exploring to reduce the size of the [Gly-Gly-Ser] linker used to separate the fluorescent and RING domains to improve the FRET signal (work in progress).

Figure 10. In-cell FRET-based reporter to screen inhibitors against RING-mediated Mdm2/MdmX interaction. **A**. Cloning and co-expression of YPet-Mdm2 and CyPet-MdmX fluorescent constructs on E. coli. YPet-Mdm2 and CyPet-MdmX were cloned into a pRSF duet expression vector. The resulting vector was used to transform E. coli cells and both proteins were expressed at room temperature for 18 h. Purified proteins were analyzed by SDS-PAGE. B. Fluorescence spectra of live *E. coli* cells expressing YPet-Mdm2 / CyPet-MdmX (FRET-on), YPet-Mdm2 (C464A) / CyPet-MdmX (FRET-off), and YPet-BARD-I / CyPet-MdmX (FRET-off). Excitation to quantify FRET signal was performed at 414 nm. Inset: quantification of fluorescent protein YPet in live E. coli cells expressing FRET-reporter indicates that the differences observed in FRET signal are not due to different expression levels of the YPet fusion protein. Cells were excited at 490 nm for YPet quatification. C. Analysis by FACS of live *E. coli* cells expressing YPet-Mdm2 / CyPet-MdmX (FRET-on); and YPet-Mdm2 (C464A) / CyPet-MdmX (FRET-off). The RING domains of Mdm2 (C464A) and BARD-I do not interact with the RING domain of MdmX and were used as negative controls to evaluate the background fluorescence of the FRET-off state.



3) Cellular uptake of MCoTI-cyclotides.

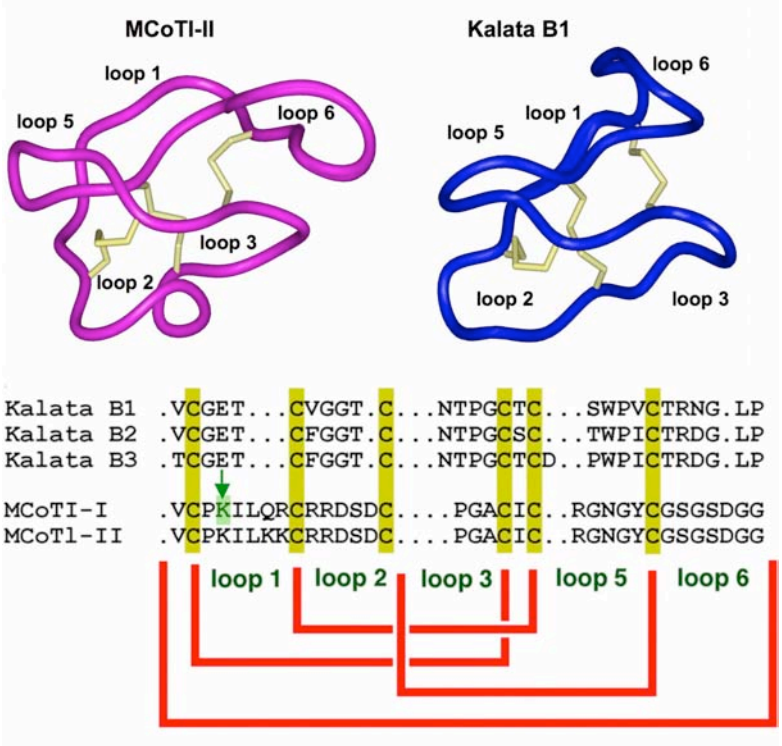
MCoTI-cyclotides have been shown recently to be able to enter human macrophages and breast cancer cell lines [11]. Internalization into macrophages was shown to be mediated mainly through macropinocytosis, a form of endocytosis that is actin-mediated and results in formation of large vesicles termed macropinosomes [12, 13]. It should be noted, however, that in this study the visualization of MCoTI-II uptake was done in fixed and not in live cells. Analysis of live cells provides the ability to visualize events in real time without the possible complications of fixation artifacts that have confounded interpretations of the uptake of Tat and other related peptides for instance [14, 15]. As well, macropinocytosis is a dominant mechanism for endocytic uptake in macrophages [16-18], unlike other cells that are not specialized for large scale sampling of extracellular fluid

and which use multiple alternative endocytic mechanisms. These mechanisms can include clathrin-mediated endocytosis, caveolar endocytosis, macropinocytosis, phagocytosis, flotillin-dependent endocytosis, as well as multiple other as yet under-characterized mechanisms [19, 20]. Intrigued by these results, we explored the cellular uptake of site-specific fluorescent-labeled MCoTI-cyclotides and studied the cellular uptake mechanisms in HeLa cells using live cell imaging by confocal fluorescence microscopy.

We have reported for the first time the cellular uptake of MCoTI-cyclotides monitored by real time confocal fluorescence microscopy imaging in live HeLa cells (manuscript under revision, Appendix: paper #1). Our results clearly show that HeLa cells readily internalize fluorescently-labeled-MCoTI-I. We found that this process is temperature-dependent and can be reversibly inhibited at 4°C, which indicates an active mechanism of internalization. The internalized cyclotide also seems to colocalize in live cells with multiple endocytic markers including, to the greatest extent, the fluid-phase endocytic marker dextran (10 KDa dextran,10K-Dex). Internalized MCoTI-I was colocalized to a lesser extent with the cholesterol/lipid dependent endocytic marker cholera toxin B (CTX-B) and the clathrin-mediated endocytic marker, EGF. Internalized MCoTI-I was localized within a fairly rapid time course with late endosomal and lysosomal compartments which engaged in rapid and directed movements suggestive of cytoskeletal involvement. MCoTI-I uptake in HeLa cells was not impaired by Latrunculin B (Lat B), a well-known inhibitor of macropinocytosis. Altogether, these data seem to indicate that MCoTI-I cyclotide is capable of internalization in live cells through multiple endocytic pathways that may be dominant in the particular cell type under study. The lack of strong preference for MCoTI-I internalization via a specific cellular internalization pathway is of significant value since the lack of endogenous affinity for a particular pathway can enable the ready re-targeting by introduction of targeting peptides within the scaffold that may enable specific and targeted endocytic uptake to a particular target cell. At the same time, the ready uptake of MCoTI-I by multiple pathways suggests accessibility, in the untargeted form, to essentially all cells.

In order to study the cellular uptake of MCoTI-cyclotides, we decided to use MCoTI-I. MCoTI-I contains only one Lys residue located in loop 1 versus MCoTI-II, which contains three Lys residues in the same loop (Fig. 11). The presence of only one Lys residue facilitates the site-specific introduction of a unique fluorophore on the sequence thus minimizing any affect that the introduction of this group may have on the cellular uptake properties of the cyclotide (see paper #1 in the Appendix section for a detailed description of the materials and methods)

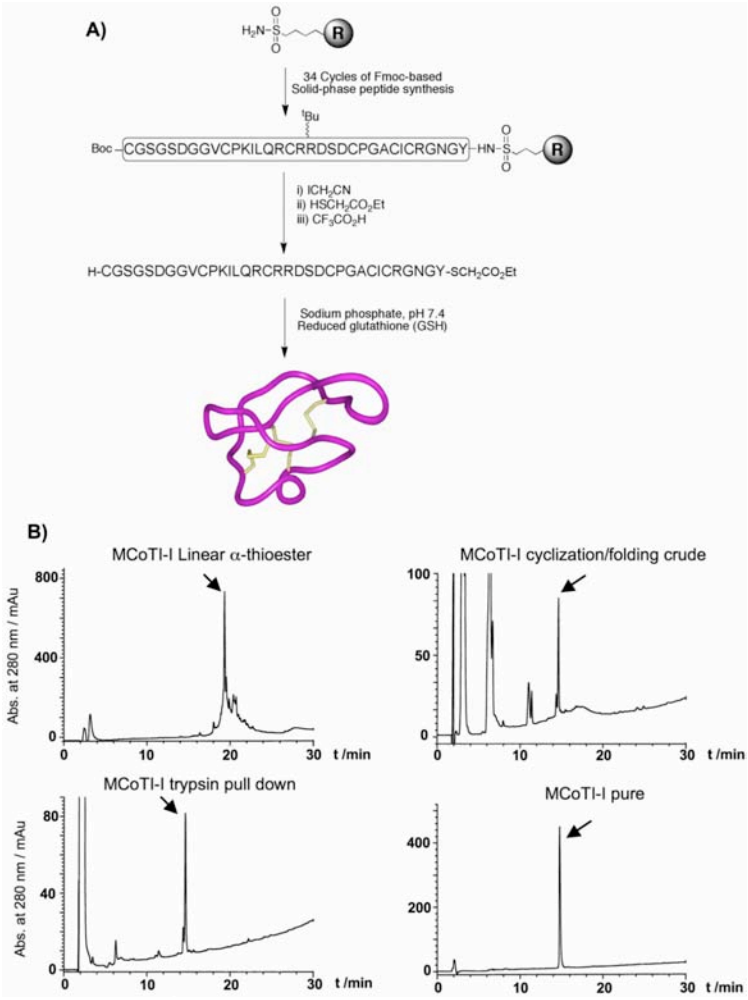
Figure 11. Primary and tertiary structures of MCoTI and kalata cyclotides. The structures of MCoTI-II (pdb ID: 1IB9 [1] and kalata B1 (pdb ID: 1NB1 [5]) are shown. Conserved cysteine residues and disulfide bonds are shown in yellow. An arrow marks residue Lys[4] located at loop 1 in MCoTI-cyclotides. This residue in MCoTI-I was used for the site-specific conjugation of AlexaFluor488 N-hydroxysuccinimide ester (AF488-OSu) through an stable amide bond.



Folded MCoTI-I cyclotide was produced either by recombinant or synthetic methods. In both cases the backbone cyclization was performed by an intramolecular native chemical ligation (NCL) [21-24] using the native Cys located to the beginning of loop 6 to facilitate the cyclization. This ligation site has been shown to give very good cyclization yields [8, 25]. Intramolecular NCL requires the presence of an N-terminal Cys residue and C-terminal α-thioester group in the same linear precursor [23, 26]. In the biosynthetic approach, the MCoTI-I linear precursor was fused in frame at their C- and N-terminus to a modified Mxe Gyrase A intein and a Met residue, respectively and expressed in *Escherichia coli* [9]. This allows the generation of the

required C-terminal thioester and N-terminal Cys residue after *in vivo* processing by endogenous Met aminopeptidase (MAP) [27, 28]. Cyclization and folding can be accomplished very efficiently *in vitro* by incubating the MCoTI-I intein fusion construct in sodium phosphate buffer at pH 7.4 in the presence of reduced glutathione (GSH). Biosynthetic MCoTI-cyclotides generated this way have been shown to adopt a native folded structure by NMR and trypsin inhibitory assays [10, 25, 28].

Figure 12. Chemical synthesis of MCoTI-I (A) Synthetic scheme used for the chemical synthesis of cyclotide MCoTI-I by Fmoc-based solid-phase peptide synthesis (B) Analytical reverse-phase HPLC traces of MCoTI-I linear precursor α-thioester, cyclization/folding crude and purified MCoTI-I by either affinity chromatograpgy using trypsin-immobilized Sepharose beads or semipreparative reverse-phase HPLC. HPLC analysis was performed in all the cases using a linear gradient of 0% to 70% buffer B over 30 min. Detection was carried out at 220 nm. An arrow indicated the desired product in each case.



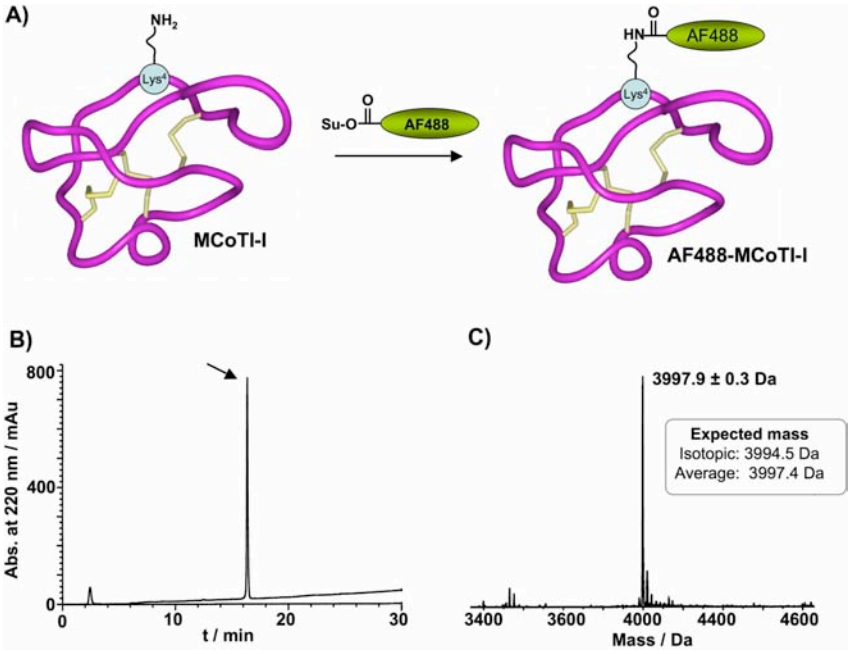
Natively folded MCoTI-II has been already successfully produced using Fmoc-based solid-phase peptide synthesis [29, 30]. Encouraged by these results we also explored the production of MCoTI-I by chemical synthesis (Fig. 12). For this purpose the MCoTI-I linear precursor α-thioester was assembled by Fmoc-based solid-phase peptide synthesis on a sulfonamide resin [31, 32] (Fig. 12A). Activation of the sulfonamide linker with iodoacetonitrile followed by cleavage with ethyl mercaptoacetate and acidolytic deprotection with TFA provided the fully protected linear peptide α-thioester (Fig. 12B). The synthetic linear precursor thioester was then efficiently cyclized and folded in one-pot reaction using sodium phosphate buffer at pH 7.5 in the presence of 2 mM GSH. The reaction was complete in 18 h and the folded product was purified by reverse-phase HPLC and characterized by ES-MS. The expected mass for folded MCoTI-I was in agreement with a folded structure (Expected mass = 3480.9 Da; measured = 3481.0 ± 0.4 Da). Synthetic folded MCoTI-II was also shown to co-elute by HPLC with recombinant natively folded MCoTI-I (data not shown). The biological activity of synthetic MCoTI-I was assayed by using a trypsin pull-down experiment [9, 25]. As shown in Figure 2B, synthetic folded MCoTI-I was specifically captured from a cyclization/folding crude reaction by trypsin-immobilized Sepharose beads [8, 9, 25], thus indicating that was adopting a native cyclotide fold.

Purified MCoTI-I was site-specifically labeled with AlexaFluor 488 (AF488) for live confocal imaging. The α-amino group of Lys^4 residue located in loop 1 was conjugated to AF488-NHS in sodium phosphate buffer at pH 7.5 for 2 h (Fig. 13A). Under these conditions the main product of the reaction was mono-labeled AF488-MCoTI as characterized by HPLC and ES-MS (expected average mass = 3997.9 Da; measured = 3997.4 ± 0.3 Da) (Figs 13C). AF488-labeled MCoTI-I was then purified by reverse-phase HPLC to remove any trace of unreacted materials (Fig. 13B).

In order to infer the correct conclusions regarding data obtained on the cellular uptake of native MCoTI-I when using modified cyclotides, like AF488-MCoTI-I for example, it is critical to be sure that they still adopt structures similar to that of the native form. MCoTI-cyclotides are extremely stable to chemical and thermal denaturation, and they have been shown to be able to withstand procedures like reverse-phase chromatography in the presence of organic solvents under acidic conditions without affecting their tertiary

structure [8, 10, 28-30, 33]. It is also unlikely that the acylation of the α-amino group of Lys^4 in MCoTI-I may disrupt the tertiary structure of this cyclotide. Craik and co-workers have previously shown that biotinylation of the three Lys residues located in loop 1 in MCoTI-II, including Lys^4 (Fig. 11) does not disrupt the native cyclotide fold of this cyclotide as determined by ^{1}H-NMR [11]. We have also recently shown that mutation of residue Lys^4 by Ala does not seem to affect the ability of this mutant to adopt a native cyclotide fold, thus indicating that the presence of positive charge residue in this position is not critical for the tertiary structure of MCoTI-I [25]. Similar findings have been also found by Leatherbarrow and coworkers, where mutation of this residue by Phe or Val was still able to render MCoTI-cyclotides able to fold correctly and have inhibitory activity against chymotrypsin and human elastase, respectively [29]. Altogether these facts suggest that residue Lys^4 is not critical for adopting the native cyclotide fold or disturbing the tertiary structure of MCoTI-cyclotides.

Figure 13. Site-specific labeling of MCoTI-I with AlexaFluor-488 N-hydroxysuccinimide ester (AF488-OSu). (A) Scheme depicting the bioconjugation process and localization of the fluorescent probe at residue Lys^4 in loop 1. (B) Analytical reverse-phase HPLC trace of pure AF488-MCoTI-I. HPLC analysis was performed using a linear gradient of 0% to 70% buffer B over 30 min. Detection was carried at 220 nm (C) ES-MS spectra of pure AF488-MCoTI-I.



To study the cellular uptake of AF488-MCoTI-I we used HeLa cells. The internalization studies were all carried out with 25 μM AF488-MCoTI-I. This concentration provided a good signal/noise ratio for live cell confocal fluorescence microscopy studies and did not show any cytotoxic effect on HeLa cells. This is in agreement with the cellular tolerance of wild-type and biotinylated MCoTI-II reported for other types of human cell lines [11]. First, we analyzed the time course of changes in cellular distribution following uptake of 25 μM AF488-MCoTI-I by incubating with the cyclotide for 1 hr and then analyzing its distribution after 1, 2, 4, 8 and 10 h. As shown in Figure 4, the internalized cyclotide was clearly visible within perinuclear punctate spots inside the cells after 1 h incubation. Observation of cells pulsed with AF488-MCoTI-I for one hour and then incubated for longer periods of time in the absence of cyclotide did not show any evidence for decreased intracellular fluorescence, while the largely perinuclear distribution of internalized MCoTI-I appeared comparable at all time points. Similar results have been also been recently reported on the internalization of biotinylated MCoTI-II in macrophage and breast cancer cell lines [11], these studies however, used fixed cells to visualize the internalized cyclotide.

In order to study the mechanism of internalization of AF488-MCoTI-I in live HeLa cells, we first explored the effect of temperature on the uptake process. Active and energy-dependent endocytic mechanisms of internalization are inhibited at 4° C [34]. The internalization of AF488-MCoTI-I was totally inhibited after a 1 h incubation at 4°C (Fig. 15). This inhibition was completely reversible and when the same cells were incubated again at 37°C for 1 h, the punctate intracellular fluorescence labeling pattern was restored. This result confirmed that the uptake of AF488-MCoTI-I in HeLa cells follows a temperature dependent active endocytic internalization pathway. It should be noted that no significant surface binding was detected at 4°C, suggesting that MCoTI-I does not bind a surface receptor, even nonspecifically. This is in agreement with studies on the MCoTI-II in fixed cells so both the MCoTI-I and MCoTI-II appear to lack specific affinity for proteins or lipids in cell membranes, unlike the kalata B1 cyclotide which shows membrane affinity [11]. This lack of endogenous affinity for a specific surface receptor or membrane constituent makes MCoTI-I ideal for engineering using more specific, receptor-directed, peptide-based, internalization motifs, within the scaffold, that might enable members of this family to have targeting enhanced to a specific cell type.

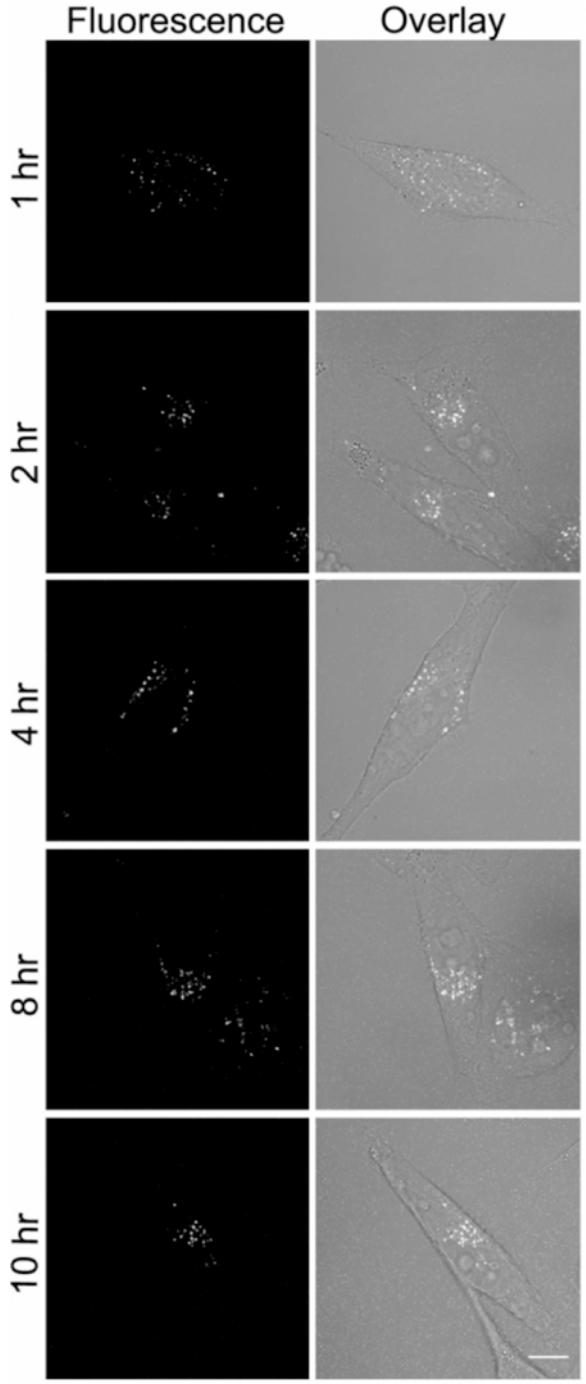
Next, we investigated the internalization pathway used by labeled-MCoTI-I to enter HeLa cells. There are several known and well-characterized mechanisms of endocytosis [35]. It is also now well established that

almost all cell-penetrating peptides (CPPs) use a combination of different endocytic pathways rather than a single endocytic mechanism [35]. A recent study showed that several CPPs (including Antennapedia/penetratin, nona-Arg and Tat peptides) can be internalized into cells by multiple endocytic pathways including macropinocytosis, clathrin-mediated endocytosis, and caveolae/lipid raft mediated endocytosis [36]. To investigate if that was the case with the internalization of AF488-MCoTI-I in HeLa cells,

we decided to look at its colocalization with various endocytic markers (Fig. 16). 10K-Dex has previously been used as a marker of fluid-phase endocytosis [12, 37-39]. CTX-B has been used as a marker for various lipid-dependent endocytic pathways [20, 40], while EGF has traditionally been a marker of clathrin-mediated endocytosis [41-43]. As shown in Figure 6, colocalization studies showed that after 1 h, AF488-MCoTI-I fluorescence was significantly colocalized with the fluorescence associated with 10K-Dex (59 ± 4% of total cyclotide fluorescence pixels were colocalized with 10K-Dex fluorescent pixels). Less colocalization was observed with fluorescent CTX-B (39 ± 4 %) and fluorescent EGF (21 ± 2 %). This data seems to suggest that AF488-MCoTI-I is primarily entering cells through fluid-phase endocytosis. The observed traces of colocalization with CTX-B and EGF also suggest that AF488-MCoTI-I could be using alternative or additional endocytic pathways. The colocalization results could also be attributed, however, to the merging of endosomal uptake vesicles generated by different pathways at the level of an early endosome. To address whether the major uptake and colocalization of AF488-MCoTI-I with 10K-Dex was due to cointernalization by macropinocytosis, we explored the inhibition of AF488-MCoTI-I uptake by Lat B, a potent inhibitor of actin polymerization, which is an essential element of macropinocytlsis [44-47]. As shown in Figure 7, Lat B did not significantly inhibit uptake of AF488-MCoTI-I (Fig. 17A) nor of 10K-Dex (data not shown). Treatment of HeLa cells with this agent resulted in a total disruption of the actin filament network (Fig. 17B). These data suggest that macropinocytosis is not responsible for uptake of either 10K-Dex nor AF488-MCoTI-I in HeLa cells.

As an extension of these inhibition studies, cells were also treated with MBCD, a well-established cholesterol-depleting agent employed for studying the involvement of lipid rafts/caveolae in endocytosis [48, 49]. Preliminary studies with MBCD suggested no significant inhibition of AF488-MCoTI-I (data not shown). Since the extent of total colocalization of AF488-MCoTI-I with CTX-B was less than 40%, it is unsurprising that no marked effect was seen by live cell microscopy. Taken together, these results seem to suggest that the uptake of AF488-MCoTI-I in HeLa cells is following multiple endocytic pathways, which is in agreement with what has been recently reported for different CPPs [36].

Figure 14. MCoTI-I distribution in HeLa cells. HeLa cells were incubated with 25 µM MCoTI-I for 1 hour, cyclotide was removed with gentle rinsing in PBS and then the cells were monitored for distribution of intracellular fluorescence at intervals from 1- 10 hours using confocal fluorescence microscopy. Bar = 10µm.



Next we explored the fate of the endocytic vesicles containing labeled MCoTI-I. There are at least two pathways that involve the cellular trafficking of endosomal vesicles. The degradative pathway includes routing

Figure 15. Endocytosis of MCoTI-I is temperature-dependent. HeLa cells were incubated with 25 μM MCoTI-I for 1 hr at 4°C. After removal of the MCoTI-I-containing media, and a gentle PBS wash, the cells were imaged. Following imaging, the MCoTI-I-containing media was replaced and the cells incubated at 37°C for 1 hr and imaged again. Bar = 10 μm.

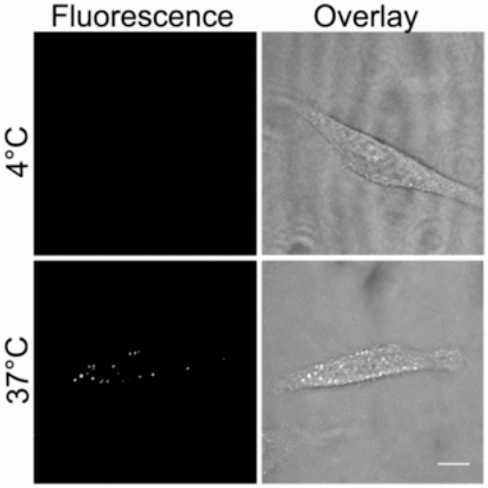
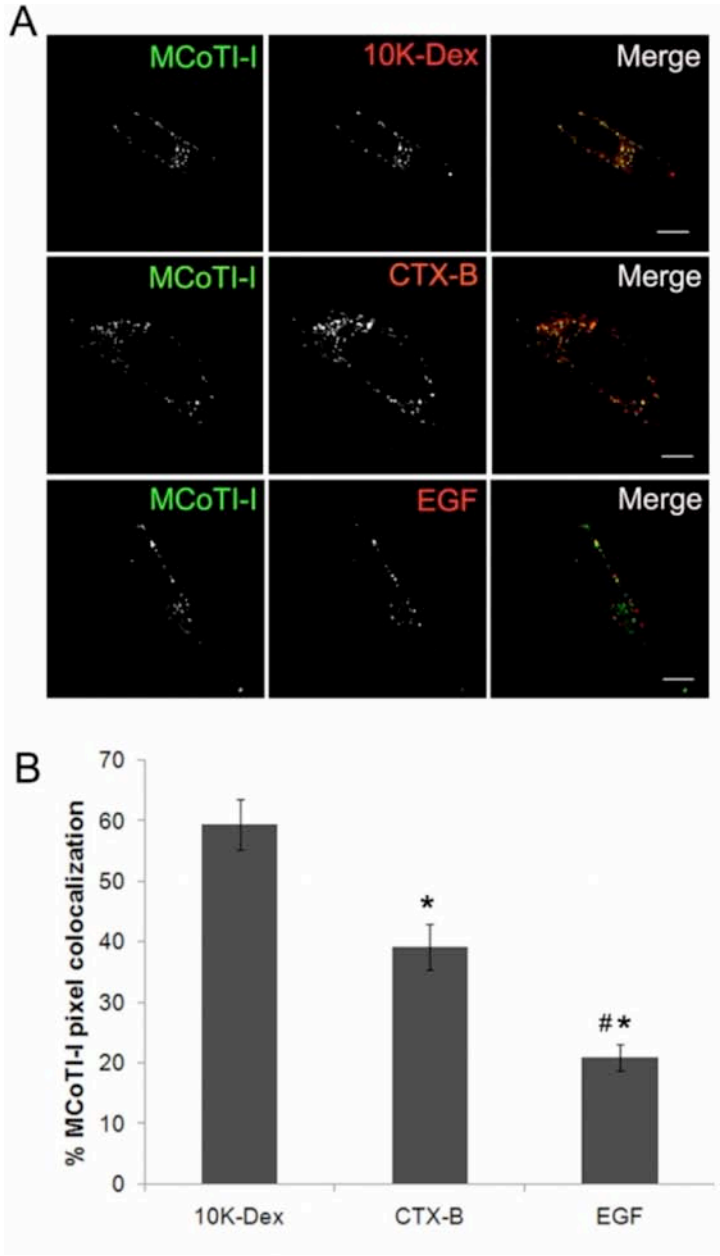


Figure 16. Colocalization of MCoTI-I with markers of endocytosis. (A) HeLa cells were incubated with 25 μM MCoTI-I and either 1 mg/ml 10K-dextran (10K-Dex), 10 μg/ml cholera toxin B (CTX-B), or 400 ng/ml epidermal growth factor (EGF) for 1 hour at 37°C as described in Materials and Methods and then imaged. Bar = 10 μm. (B) Quantification of pixel colocalization was done using the Zeiss LSM software for image analysis and measures the % of total fluorescent AF488 MCoTI-I pixels in the ROI relative to red pixels associated with different endocytic markers. (n = 13 cells for 10K-Dex, n = 11 cells for CTX-B and n = 10 cells for EGF, with cells assessed across 3 different experiments, * $p \leq 0.05$ relative to 10K-Dex, # $p \leq 0.05$ relative to CTX-B).



of internalized materials from early endosomes via late endosomes to lysosomes where degradation of internalized materials occurs within the cells. On the other hand, recycling endosomes sort material internalized into early endosomes and are responsible for effluxing internalized material back to the cellular membrane [50]. If labeled-MCoTI-I was localized in recycling endosomes, it would be expected that its concentration in the cell would decrease and/or accumulate on the membrane over time, which was not the case in the time course experiment following the cellular fate of internalized cyclotide (Fig. 14). To explore the potential localization of labeled-MCoTI-I in lysosomes we first used LysoTracker Red (LysoRed). This pH sensitive fluorescent probe is utilized for identifying acidic organelles, such as lysosomes and late endosomes, in live cells. As shown in Figure 8A, significant colocalization (60 ± 4.0% as determined by pixel colocalization analysis) of LysoRed and AF488-MCoTI-I was observed after treating the cells for 1h with both agents. As an extension of these experiments, we also investigated the colocalization of labeled-MCoTI-I and lysosomal-associated membrane protein 1 (Lamp1), an established mature lysosomal marker [51, 52]. For this experiment, live HeLa cells were first infected with a Red Fluorescent Protein (RFP)-Lamp1-expressing BacMam virus. The next day the cells were incubated with AF488-MCoTI-I for 1 h and imaged. As shown in Figure 8B, colocalization was also seen for AF488-MCoTI-I and RFP-Lamp1 (38 ± 5%, as determined by pixel colocalization analysis), suggesting that even after 1 h, significant MCoTI-I has already reached the lysosomal compartments. Our data suggest that after 1 h, a significant amount of MCoTI-I (≈40%) has trafficked through the endosomal pathway to the

lysosomes and that ≈20% is already localized in late endosomes or other types of acidic organelles. It has previously been reported that the perinuclear steady-state distribution of lysosomes is a balance between movement on microtubules and actin filaments [53-55]. Likewise, movement from early endosomal compartments to late endosomes to lysosomes has also been shown to rely on the microtubule network [56, 57]. As an extension of these experiments, and to investigate whether MCoTI-I-containing vesicles were actively trafficking inside the cell, we captured time-lapse video of cells after incubation with MCoTI-I for 1 h. Indeed, the time-lapse capture showed active movements of MCoTI-I-containing vesicles (Fig. 19). Directed short- and long-range movements could be seen, characteristic of movement on cytoskeletal filaments. These results suggest that while a large portion of MCoTI-I has reached lysosomal compartments by 1 h, and some of the movements seen may be attributed to the steady-state distribution of lysosomes, the remaining cyclotide may still be trafficking through the cell from other membrane compartments, likely within late endosomes.

Figure 17. Disruption of actin does not inhibit MCoTI-I uptake. (A) HeLa cells were untreated (control) or treated with Lat B (2 µM) for 30 min at 37°C prior to addition of 25 µM MCoTI-I. Following uptake for 1 hr at 37°C, the cells were imaged using confocal fluorescence microscopy. Bar = 10 µm. (B) HeLa cells without treatment (control) or treated with 2 µM Lat B for 30 min at 37°C were fixed and labeled with rhodamine—phalloidin to label actin (red) and DAPI to label nuclei (blue). Bar = 10 µm.

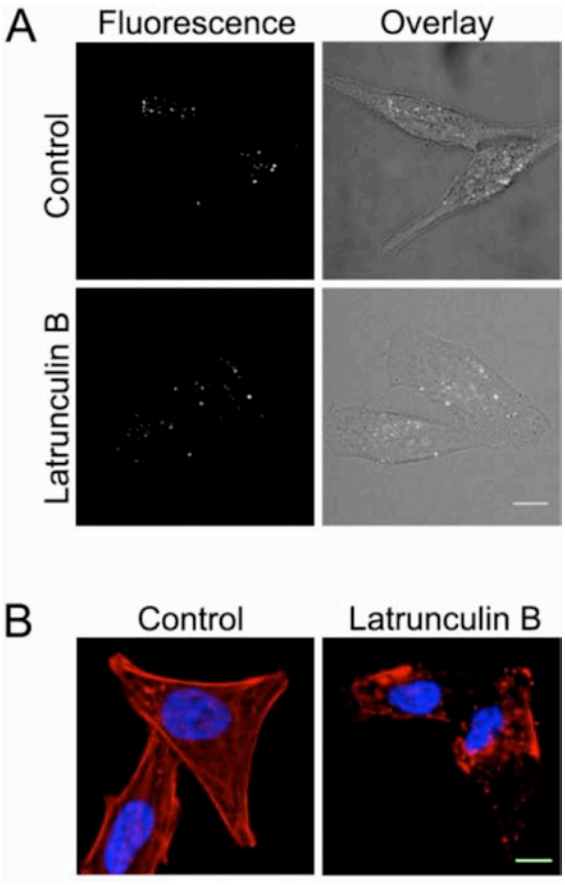
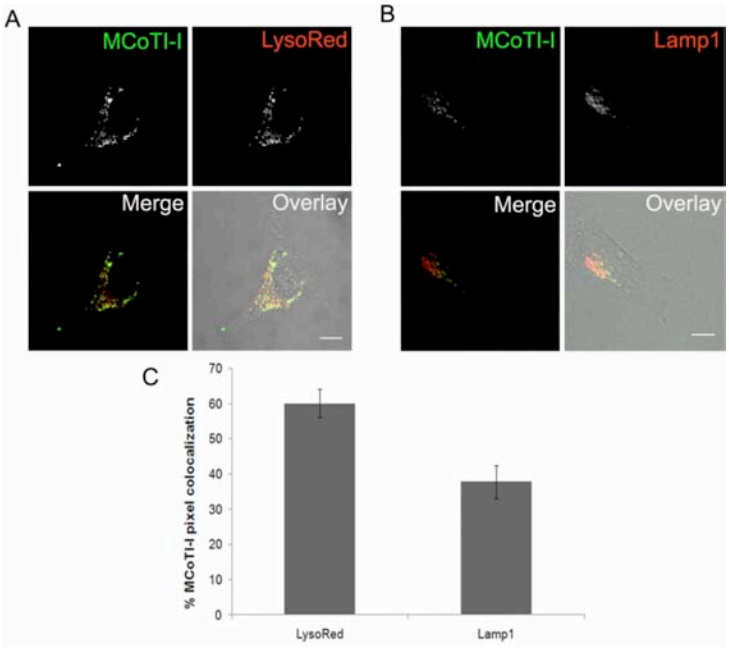


Figure 18. MCoTI-I is colocalized with lysosomal compartments. (A) Untreated or BacMam-RFP-Lamp1 treated HeLa cells were incubated with 25 µM MCoTI-I and LysoTracker Red, or MCoTI-I alone, for 1 hr at 37°C as described in Materials and Methods and then imaged. Bar = 10 µm. (B) Quantification of % of total fluorescent AF488-MCoTI-I pixel colocalization with fluorescent pixels associated with both markers was done using the Zeiss LSM software for image analysis. (*n* = 14 cells for LysoTracker Red and *n* = 11 cells for Lamp1 with cells selected from 3 separate experiments).

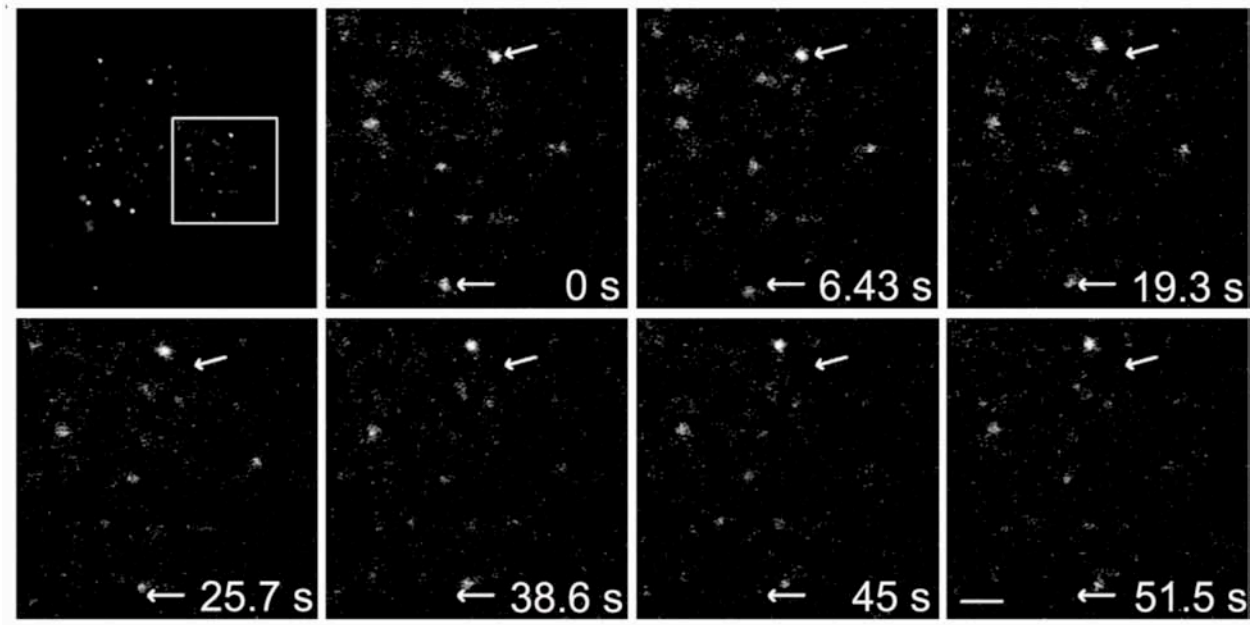


Craik and co-workers have recently reported the uptake of biotinylated-MCoTI-II by human macrophages and breast cancer cell lines [11]. This work concluded that the uptake of MCoTI-II in macrophages is mediated by macropinocytosis and that the cyclotide accumulates in macropinosomes without trafficking to the lysosome. MCoTI-II shares high homology with MCoTI-I (≈97% homology, see Fig. 11) and similar biological activity. Despite their similarities, the differences in the cellular uptake and trafficking of MCoTI-cyclotides by macrophages versus HeLa cells could be attributed to the cellular differences in endocytic preferences for these two very different cell types. Macrophage cells are specialized in large scale sampling of extracellular fluid using macropiniocytosis as the dominant endocytic pathway. Meanwhile, other types of cells may use multiple endocytic pathways as has been recently shown for the uptake of different CPPs in HeLa cells [36].

At this point we cannot be certain if some labeled MCoTI-I is able to escape from endosomal/lysosomal compartments into the cytosol. The ability to track the release of fluorescent-labeled molecules from cellular

vesicles is limited using live cell imaging of fluorescence signal primarily due to the large dilution effect if the molecule is able to escape the highly confined volume of the vesicle into the larger cytosolic volume. One way to demonstrate the release of peptide into the cytosol, however, would be by using labels with better detection sensitivity or incorporating a biological activity that can be measured in the cellular cytosol. For example, the presence of Tyr residues in both MCoTI-cyclotides should facilitate the incorporation of radioactive iodine into the phenolic ring of Tyr with minimal disruption of the native structure of the cyclotide. The incorporation or grafting of biological peptides into the MCoTI scaffold could also provide proof of endosomal/lysosomal escape if such biological activity could be measured only in the cytosol. This approach has been already used to demonstrate endosomal escape of CPPs such as the TAT peptide [58, 59]. The retention of fluorescence signal in the perinuclear, lysosomal compartments for a period of up to 10 hrs suggests that most of the cyclotide remains within these compartments however; given the flexibility of the cyclotide backbone to accommodate multiple peptide sequences, subsequent studies may explore the ability of targeting and endosomolytic sequences for concomitant targeted entry and endosomal/lysosomal escape into cytosol.

Figure 19. MCoTI-I-containing vesicles are in motion. HeLa cells were incubated with 25 μM MCoTI-I for 1 hr at 37°C and then imaged using time-lapse microscopy as described in Materials and Methods. Arrows indicate position of the moving vesicle at 0 min while displacement of the fluorescent vesicle relative to the arrow shows the extent of movement over time. Bar = 2 μm.



In conclusion, this study reports on the first analysis of intracellular uptake of MCoTI-I cyclotide using live cell imaging by confocal fluorescence microscopy. Cyclotides represent a novel new platform for drug development. Their stability, conferred by the cyclic cystine knot, their small size, their amenability to both chemical and biological synthesis and their flexible tolerance to sequence variation make them ideal for grafting of biologically-active therapeutic epitopes. As we show herein, they are also capable in the unmodified state of utilizing multiple cellular endocytic pathways for internalization. Their ease of access makes them readily accessible in their current state to endosomal/lysosomal compartments of virtually any cell. Without an apparent strong preference for an existing cellular pathway nor surface-expressed epitope in HeLa cells (nor in other studies with MCoTI-II in macrophages [11]), they appear highly amenable to retargeting to exploit a particular target cell's dominant internalization pathway and/or unique surface receptor repertoire, along with the targeted introduction of biologically-active therapeutic motifs.

Key Research Accomplishments (2010-2011)

- We have reported biosynthesis of a genetically-encoded library of MCoTI-based cyclotides containing a comprehensive suite of amino acid mutants. The mutagenesis results obtained in our work highlighs the extreme robustness of the cyclotide scaffold to mutations. The results obtained are key for the design of larger combinatorial libraries to screen RING Mdm2/MdmX antagonists (Appendix Section: paper #7).

- We have accomplished the biosynthesis of a large combinatorial library (≈ 10^6 members) of genetically-encoded cyclotides in *Escherichia coli* using loop 2 of cyclotide MCoTI-I as molecular template. This is the first time that a library with such a high diversity has been produced. This library will be used to screen RING Mdm2/MdmX antagonists.

- We have accomplished the cloning of cyclotide-based libraries into different *E. coli* expression plasmids with different promoters, origin of replication and antibiotic resistance to allow the screening of the libraries using cell-based reporters encoded in orthogonal plasmids. This will allow in-cell screening of genetically encoded cyclotide-based libraries.

- We have carried out the first ever reported study of the backbone dynamics of a natively folded MCoTI-I cyclotide in the free state and complexed to its binding partner trypsin in solution. This accomplishment is critical to understand the dynamics and better define the loops that are best amenable for randomization in the generation of MCoTI-based libraries (Appendix Section: paper #5)

- We have developed a FRET-based reporter using fluorescent proteins CyPet and YPet able to detect both in vitro and in vivo protein-protein antagonists for the RING-mediated Hdm2/HdmX interaction. We are now optimizing the constructs to maximize the fluorescence signal. This accomplishment is key for the success of the project.

- We have investigated the cellular uptake of cyclotide, MCoTI-I in live HeLa cells. Using real time confocal fluorescence microscopy imaging we show that MCoTI-I is readily internalized in live HeLa cells and that its endocytosis is temperature-dependent. Endocytosis of MCoTI-I in HeLa cells is achieved primarily through fluid-phase endocytosis, as evidenced by its significant colocalization with 10K-dextran, but also through other pathways as well, as evidenced by its colocalization with markers for cholesterol-dependent and clathrin-mediated endocytosis, cholera toxin B and EGF respectively. Uptake does not appear to occur via macropinocytosis as inhibition of this pathway by Latrunculin B-induced disassembly of actin filaments did not affect MCoTI-I uptake. As well, a significant amount of MCoTI-I accumulates in late endosomal and lysosomal compartments and MCoTI-I-containing vesicles continue to exhibit directed movements. These findings demonstrate internalization of MCoTI-I through endocytic pathways that are dominant in the cell type investigated, suggesting that this cyclotide has ready access to general endosomal/lysosomal pathways but could readily be re-targeted to specific receptors through addition of targeting ligands (Appendix Section: paper #1)

- We have also successfully accomplished the chemical synthesis of wild-type and chemically modified MCoTI-cyclotides. This is key for the molecular characterization of biomolecular intereactions and for further development of cyclotide leads to improve their pharmacokinetic properties (Appendix Section: paper # 1).

Reportable Outcomes

Peer-reviewed Publications Submitted and Published:

- J. Contreras, A. Y. O. Elnagar, S. Hamm-Alvarez and **J. A. Camarero** (2011) Cellular Uptake of cyclotide MCoTI-I follows multiple endocytic pathways, *J. Control Release*, under revision (Appendix: paper #1).

- **J. A. Camarero** (2011) Legume cyclotides shed new light on the genetic origin of knotted circular proteins, *Proc. Natl. Acad. Sci. USA*, in press (Appendix: paper #2).

- L. Berrade, Angie E. Garcia **and J. A. Camarero** (2010) Protein Microarrays: Novel Developments and Applications, *Pharmacol. Res*., DOI 10.1007/s11095-010-0325-1 (Appendix: paper #3).

- A. E. Garcia and **J. A. Camarero** (2010) Biological Activities of Natural and Engineered Cyclotides, a Novel Molecular Scaffold for Peptide-Based Therapeutics, *Curr. Mol. Pharmacol*., **3**(3), 153-163 (Appendix: paper #4)

- S. S. Puttamadappa, K. Jagadish, A. Shekhtman and **J. A. Camarero** (2010) Backbone dynamics of cyclotide MCoTI-I free and complexed with trypsin, *Angew. Chem. Int. Ed.*, **49**(39), 7030-7034 (Appendix: paper #5)

- K. Jagadish and **J. A. Camarero** (2010) Cyclotides, a promising molecular scaffold for peptide-based therapeutics, *Biopolymers*, **94**(5), 611-616 (Paper #6).

- J. Austin, Wan Wang, Swamy Puttamadappa, Alexander Shekhtman and **J. A. Camarero** (2010) Biosynthesis and biological screening of a genetically-encoded library based on the cyclotide MCoTI-I, *ChemBioChem,* **10**(16), 2663-2670 (Paper #7).

Oral presentations:

- Seventh Annual PEGS (Protein Engineering Summit) Conference, Phage and Yeast Display of Antibodies and Proteins Session, May 9-13, 2011, Boston.

- 2011 Spring ACS National Meeting, Division of Biological, Invited Lecture to the Ralph F. Hirschmann Award in Peptide Chemistry: Symposium in Honor of David J. Craik, Anaheim, March 28, 2011, Anaheim.

- University of Uppsala, Invited seminar to The Svedberg Lecture Series, March 24, 2011, Uppsala, Sweden.

- Pacifichem 2010: International Chemical Congress of Pacific Basic Societies, December 15-20, Honolulu, Hawaii.

- Roche Colorado Corporation Peptide Symposium (RCCPS) 2010, Cyclotides, a novel natural peptide scaffold for drug discovery, September 14-16, Boulder, Colorado.

- Natural Peptides to Drugs (NP2D) 4th International Congress, April 11-14, 2010, Zermatt, Switzerland.

Patent Applications

- Composition and methods for the rapid biosynthesis and in vivo screening of biologically relevant peptides, J. A. Camarero, PCT International Application No. PCT/US2010/039720.

- Novel cyclotide-based polypeptides with antiviral and anticancer activity, J. A. Camarero and J. Jung, filed US patent application #61/283,096.

Conclusion

The results accomplished during the first 12 months of the proposal are extremely encouraging. We have shown that protein splicing can be used for the generation of large libraries of genetically encoded libraries (≈10^6 members), and in principle large libraries can be generated by randomizing 2 loops (≈10^9 members) (***Aim #1, Task 2, accomplished***).

We have also recently reported the backbone dynamics of the cyclotide MCoTI-I in the free state and complexed to its binding partner trypsin in solution [3]. This is the first time the backbone dynamics of a natively folded cyclotide has been reported in the literature. Such insight will help us in the design of optimal focused libraries than can be used for the discovery of new cyclotides sequences with novel biological activities. We have also shown that MCoTI-cyclotides can be readily synthesized by chemical means thus allowing the introduction of non-natural amino acids. More importantly, we have designed a FRET-based fluorescence reporter that can be used for monitor inhibition of the RING-mediated Mdm2/MdmX interaction inside living cells using high throughput cell-sorting techniques (***Aim #1, Task 1 and milestone #1 accomplished***). Interfacing the production of cyclotide-based cell-libraries with in-cell screening methods will allow the rapid selection of a novel type of ultrastable specific inhibitors for targeting Mdm2/MdmX interaction.

Next year, we will create more libraries using different loops of the cyclotide MCoTI-I. We will also prepare larger libraries with 2 loops randomized. These libraries will be screened using fluorescence-activated cell-sorting using our FRET-based genetically encoded reporter for the RING-mediated Mdm2/MdmX interaction. We will also continue studying the ability of some cyclotides to cross cellular membranes using radiolabeling and confocal microscopy. The study will be accomplished in several cell lines including prostate cancer cell lines.

References

[1] Felizmenio-Quimio, M.E.;Daly, N.L.;Craik, D.J. Circular proteins in plants: solution structure of a novel macrocyclic trypsin inhibitor from Momordica cochinchinensis. *J Biol Chem*, **2001**, *276*, 22875-22882.

[2] Zhang, Z.;Li, M.;Wang, H.;Agrawal, S.;Zhang, R. Antisense therapy targeting MDM2 oncogene in prostate cancer: Effects on proliferation, apoptosis, multiple gene expression, and chemotherapy. *Proc Natl Acad Sci U S A*, **2003**, *100*, 11636-11641.

[3] Logan, I.R.;McNeill, H.V.;Cook, S.;Lu, X.;Lunec, J.;Robson, C.N. Analysis of the MDM2 antagonist nutlin-3 in human prostate cancer cells. *Prostate*, **2007**, *67*, 900-906.

[4] Vassilev, L.T.;Vu, B.T.;Graves, B.;Carvajal, D.;Podlaski, F.;Filipovic, Z.;Kong, N.;Kammlott, U.;Lukacs, C.;Klein, C.;Fotouhi, N.;Liu, E.A. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science*, **2004**, *303*, 844-848.

[5] Rosengren, K.J.;Daly, N.L.;Plan, M.R.;Waine, C.;Craik, D.J. Twists, knots, and rings in proteins. Structural definition of the cyclotide framework. *J Biol Chem*, **2003**, *278*, 8606-8616.

[6] Vousden, K.H.;Lane, D.P. p53 in health and disease. *Nat Rev Mol Cell Biol*, **2007**, *8*, 275-283.

[7] Kimura, R.H.;Steenblock, E.R.;Camarero, J.A. Development of a cell-based fluorescence resonance energy transfer reporter for Bacillus anthracis lethal factor protease. *Anal Biochem*, **2007**, *369*, 60-70.

[8] Camarero, J.A.;Kimura, R.H.;Woo, Y.H.;Shekhtman, A.;Cantor, J. Biosynthesis of a fully functional cyclotide inside living bacterial cells. *Chembiochem*, **2007**, *8*, 1363-1366.

[9] Austin, J.;Kimura, R.H.;Woo, Y.H.;Camarero, J.A. In vivo biosynthesis of an Ala-scan library based on the cyclic peptide SFTI-1. *Amino Acids*, **2010**, *38*, 1313-1322.

[10] Puttamadappa, S.S.;Jagadish, K.;Shekhtman, A.;Camarero, J.A. Backbone Dynamics of Cyclotide MCoTI-I Free and Complexed with Trypsin. *Angew Chem Int Ed Engl*, **2010**,

[11] Greenwood, K.P.;Daly, N.L.;Brown, D.L.;Stow, J.L.;Craik, D.J. The cyclic cystine knot miniprotein MCoTI-II is internalized into cells by macropinocytosis. *Int J Biochem Cell Biol*, **2007**, *39*, 2252-2264.

[12] Hewlett, L.J.;Prescott, A.R.;Watts, C. The coated pit and macropinocytic pathways serve distinct endosome populations. *J Cell Biol*, **1994**, *124*, 689-703.

[13] Kerr, M.C.;Teasdale, R.D. Defining macropinocytosis. *Traffic*, **2009**, *10*, 364-371.

[14] Lundberg, M.;Johansson, M. Positively charged DNA-binding proteins cause apparent cell membrane translocation. *Biochem Biophys Res Commun*, **2002**, *291*, 367-371.

[15] Richard, J.P.;Melikov, K.;Vives, E.;Ramos, C.;Verbeure, B.;Gait, M.J.;Chernomordik, L.V.;Lebleu, B. Cell-penetrating peptides. A reevaluation of the mechanism of cellular uptake. *J Biol Chem*, **2003**, *278*, 585-590.

[16] Lim, J.P.;Gleeson, P.A. Macropinocytosis: an endocytic pathway for internalising large gulps. *Immunol Cell Biol*, **2011**,

[17] Norbury, C.C.;Hewlett, L.J.;Prescott, A.R.;Shastri, N.;Watts, C. Class I MHC presentation of exogenous soluble antigen via macropinocytosis in bone marrow macrophages. *Immunity*, **1995**, *3*, 783-791.

[18] Steinman, R.M.;Brodie, S.E.;Cohn, Z.A. Membrane flow during pinocytosis. A stereologic analysis. *J Cell Biol*, **1976**, *68*, 665-687.

[19] Conner, S.D.;Schmid, S.L. Regulated portals of entry into the cell. *Nature*, **2003**, *422*, 37-44.

[20] Doherty, G.J.;McMahon, H.T. Mechanisms of endocytosis. *Annu Rev Biochem*, **2009**, *78*, 857-902.

[21] Camarero, J.A.;Muir, T.W. Chemoselective backbone cyclization of unprotected peptides. *J. Chem. Soc., Chem. Comm.*, **1997**, *1997*, 1369-1370.

[22] Zhang, L.;Tam, J.P. Synthesis and application of unprotected cyclic peptides as building blocks for peptide dendrimers. *J. Am. Chem. Soc.*, **1997**, *119*, 2363-2370.

[23] Camarero, J.A.;Pavel, J.;Muir, T.W. Chemical Synthesis of a Circular Protein Domain: Evidence for Folding-Assisted Cyclization. *Angew. Chem. Int. Ed.*, **1998**, *37*, 347-349.

[24] Shao, Y.;Lu, W.Y.;Kent, S.B.H. A novel method to synthesize cyclic peptides. *Tetrahedron Lett.*, **1998**, *39*, 3911-3914.

[25] Austin, J.;Wang, W.;Puttamadappa, S.;Shekhtman, A.;Camarero, J.A. Biosynthesis and biological screening of a genetically encoded library based on the cyclotide MCoTI-I. *Chembiochem*, **2009**, *10*, 2663-2670.

[26] Camarero, J.A.;Muir, T.W. Biosynthesis of a Head-to-Tail Cyclized Protein with Improved Biological Activity. *J. Am. Chem. Soc.*, **1999**, *121*, 5597-5598.

[27] Camarero, J.A.;Fushman, D.;Cowburn, D.;Muir, T.W. Peptide chemical ligation inside living cells: in vivo generation of a circular protein domain. *Bioorg Med Chem*, **2001**, *9*, 2479-2484.

[28] Kimura, R.H.;Tran, A.T.;Camarero, J.A. Biosynthesis of the cyclotide kalata B1 by using protein splicing. *Angewandte Chemie (International ed*, **2006**, *45*, 973-976.

[29] Thongyoo, P.;Roque-Rosell, N.;Leatherbarrow, R.J.;Tate, E.W. Chemical and biomimetic total syntheses of natural and engineered MCoTI cyclotides. *Org Biomol Chem*, **2008**, *6*, 1462-1470.

[30] Thongyoo, P.;Tate, E.W.;Leatherbarrow, R.J. Total synthesis of the macrocyclic cysteine knot microprotein MCoTI-II. *Chem Commun (Camb)*, **2006**, 2848-2850.
[31] Ingenito, R.;Bianchi, E.;Fattori, D.;Pessi, A. Solid phase synthesis of peptide C-terminal thioesters by Fmoc/t-Bu chemistry
Source. *J. Am. Chem. Soc.*, **1999**, *121*, 11369-11374.
[32] Shin, Y.;Winans, K.A.;Backes, B.J.;Kent, S.B.H.;Ellman, J.A.;Bertozzi, C.R. Fmoc-Based Synthesis of Peptide-^{a}Thioesters: Application to the Total Chemical Synthesis of a Glycoprotein by Native Chemical Ligation. *J. Am. Chem. Soc.*, **1999**, *121*, 11684-11689.
[33] Daly, N.L.;Love, S.;Alewood, P.F.;Craik, D.J. Chemical synthesis and folding pathways of large cyclic polypeptides: studies of the cystine knot polypeptide kalata B1. *Biochemistry*, **1999**, *38*, 10606-10614.
[34] Silverstein, S.C.;Steinman, R.M.;Cohn, Z.A. Endocytosis. *Annu Rev Biochem*, **1977**, *46*, 669-722.
[35] Fonseca, S.B.;Pereira, M.P.;Kelley, S.O. Recent advances in the use of cell-penetrating peptides for medical and biological applications. *Adv Drug Deliv Rev*, **2009**, *61*, 953-964.
[36] Duchardt, F.;Fotin-Mleczek, M.;Schwarz, H.;Fischer, R.;Brock, R. A comprehensive model for the cellular uptake of cationic cell-penetrating peptides. *Traffic*, **2007**, *8*, 848-866.
[37] Brown, D.;Sabolic, I. Endosomal pathways for water channel and proton pump recycling in kidney epithelial cells. *J Cell Sci Suppl*, **1993**, *17*, 49-59.
[38] Shurety, W.;Stewart, N.L.;Stow, J.L. Fluid-phase markers in the basolateral endocytic pathway accumulate in response to the actin assembly-promoting drug Jasplakinolide. *Mol Biol Cell*, **1998**, *9*, 957-975.
[39] Thompson, K.;Rogers, M.J.;Coxon, F.P.;Crockett, J.C. Cytosolic entry of bisphosphonate drugs requires acidification of vesicles after fluid-phase endocytosis. *Mol Pharmacol*, **2006**, *69*, 1624-1632.
[40] Nabi, I.R.;Le, P.U. Caveolae/raft-dependent endocytosis. *J Cell Biol*, **2003**, *161*, 673-677.
[41] Huang, F.;Khvorova, A.;Marshall, W.;Sorkin, A. Analysis of clathrin-mediated endocytosis of epidermal growth factor receptor by RNA interference. *J Biol Chem*, **2004**, *279*, 16657-16661.
[42] Jiang, X.;Huang, F.;Marusyk, A.;Sorkin, A. Grb2 regulates internalization of EGF receptors through clathrin-coated pits. *Mol Biol Cell*, **2003**, *14*, 858-870.
[43] Sorkin, A.;Von Zastrow, M. Signal transduction and endocytosis: close encounters of many kinds. *Nat Rev Mol Cell Biol*, **2002**, *3*, 600-614.
[44] Burckhardt, C.J.;Greber, U.F. Virus movements on the plasma membrane support infection and transmission between cells. *PLoS Pathog*, **2009**, *5*, e1000621.
[45] Jerdeva, G.V.;Wu, K.;Yarber, F.A.;Rhodes, C.J.;Kalman, D.;Schechter, J.E.;Hamm-Alvarez, S.F. Actin and non-muscle myosin II facilitate apical exocytosis of tear proteins in rabbit lacrimal acinar epithelial cells. *J Cell Sci*, **2005**, *118*, 4797-4812.
[46] Mercer, J.;Helenius, A. Virus entry by macropinocytosis. *Nat Cell Biol*, **2009**, *11*, 510-520.
[47] Peterson, J.R.;Mitchison, T.J. Small molecules, big impact: a history of chemical inhibitors and the cytoskeleton. *Chem Biol*, **2002**, *9*, 1275-1285.
[48] Le, P.U.;Nabi, I.R. Distinct caveolae-mediated endocytic pathways target the Golgi apparatus and the endoplasmic reticulum. *J Cell Sci*, **2003**, *116*, 1059-1071.
[49] Parton, R.G.;Simons, K. The multiple faces of caveolae. *Nat Rev Mol Cell Biol*, **2007**, *8*, 185-194.
[50] Spang, A. On the fate of early endosomes. *Biol Chem*, **2009**, *390*, 753-759.
[51] Eskelinen, E.L.;Tanaka, Y.;Saftig, P. At the acidic edge: emerging functions for lysosomal membrane proteins. *Trends Cell Biol*, **2003**, *13*, 137-145.
[52] Fukuda, M. Lysosomal membrane glycoproteins. Structure, biosynthesis, and intracellular trafficking. *J Biol Chem*, **1991**, *266*, 21327-21330.
[53] Cordonnier, M.N.;Dauzonne, D.;Louvard, D.;Coudrier, E. Actin filaments and myosin I alpha cooperate with microtubules for the movement of lysosomes. *Mol Biol Cell*, **2001**, *12*, 4013-4029.
[54] Matteoni, R.;Kreis, T.E. Translocation and clustering of endosomes and lysosomes depends on microtubules. *J Cell Biol*, **1987**, *105*, 1253-1265.
[55] Taunton, J.;Rowning, B.A.;Coughlin, M.L.;Wu, M.;Moon, R.T.;Mitchison, T.J.;Larabell, C.A. Actin-dependent propulsion of endosomes and lysosomes by recruitment of N-WASP. *J Cell Biol*, **2000**, *148*, 519-530.
[56] Aniento, F.;Emans, N.;Griffiths, G.;Gruenberg, J. Cytoplasmic dynein-dependent vesicular transport from early to late endosomes. *J Cell Biol*, **1993**, *123*, 1373-1387.
[57] Loubery, S.;Wilhelm, C.;Hurbain, I.;Neveu, S.;Louvard, D.;Coudrier, E. Different microtubule motors move early and late endocytic compartments. *Traffic*, **2008**, *9*, 492-509.
[58] Frankel, A.D.;Pabo, C.O. Cellular uptake of the tat protein from human immunodeficiency virus. *Cell*, **1988**, *55*, 1189-1193.

[59] Lee, J.S.;Li, Q.;Lee, J.Y.;Lee, S.H.;Jeong, J.H.;Lee, H.R.;Chang, H.;Zhou, F.C.;Gao, S.J.;Liang, C.;Jung, J.U. FLIP-mediated autophagy regulation in cell death control. *Nat Cell Biol*, **2009**, *11*, 1355-1362.

Appendix

Abstract for oral presentations during 2010-2011.

Cyclotides, a novel natural peptide scaffold for drug discovery

Julio A. Camarero*
University of Southern California, Los Angeles, CA 90033

Cyclotides are a new emerging family of large plant-derived backbone-cyclized polypeptides (≈28-37 amino acids long) that share a disulfide-stabilized core (3 disulfide bonds) characterized by an unusual knotted. Cyclotides contrast with other circular poylpeptides in that they have a well-defined three-dimensional structure, and despite their small size, can be considered as miniproteins. The main features of cyclotides are therefore a remarkable stability due to the cystine knot, a small size making them readily accessible to chemical synthesis, and an excellent tolerance to sequence variations. For example, the first cyclotide to be discovered, kalata B1, is an orally effective uterotonic, and other cyclotides have been shown to cross the cell membrane through macro-pinocytosis Cyclotides thus appear as promising leads or frameworks for peptide drug design.

We report for the first time the *in vivo* biosynthesis of natively-folded MCoTI-II inside live *E. coli* cells. The cyclotide MCoTI-II is a powerful trypsin inhibitor recently isolated from the seeds of *Momordica cochinchinensis*, a plant member of *cucurbitaceae* family. Biosynthesis of genetically encoded cyclotide-based libraries opens the possibility of using single cells as microfactories where the biosynthesis and screening of particular inhibitor can take place in a single process within the same cellular cytoplasm. The cyclotide scaffold has a tremendous potential for the development of therapeutic leads based on their extraordinary stability and potential for grafting applications. We will also report the design and biosynthesis of a MCoTI-grafted cyclotide with the ability to specifically induce programmed cell-death in-cell assays using different tumor cell lines.

Elsevier Editorial System(tm) for Journal of Controlled Release
Manuscript Draft

Manuscript Number: JCR-D-11-00481

Title: Cellular Uptake of cyclotide MCoTI-I follows multiple endocytic pathways

Article Type: Special Issue: 15th SLC DDS

Section/Category:

Keywords: cyclotides, cell penetrating peptides, endocytosis, peptide scaffold

Corresponding Author: Dt Julio A Camarero, PhD

Corresponding Author's Institution: University of Southern California

First Author: Julio A Camarero, PhD

Order of Authors: Julio A Camarero, PhD; Janette Contreras; Ahmed Elnagar; Sarah Hamm-Alvarez

Abstract: Cyclotides are plant-derived proteins that naturally exhibit various biological activities and whose unique cyclic structure makes them remarkably stable and resistant to denaturation or degradation. These attributes, among others, make them ideally suited for use as drug development tools. This study investigated the cellular uptake of cyclotide, MCoTI-I in live HeLa cells. Using real time confocal fluorescence microscopy imaging we show that MCoTI-I is readily internalized in live HeLa cells and that its endocytosis is temperature-dependent. Endocytosis of MCoTI-I in HeLa cells is achieved primarily through fluid-phase endocytosis, as evidenced by its significant colocalization with 10K-dextran, but also through other pathways as well, as evidenced by its colocalization with markers for cholesterol-dependent and clathrin-mediated endocytosis, cholera toxin B and EGF respectively. Uptake does not appear to occur via macropinocytosis as inhibition of this pathway by Latrunculin B-induced disassembly of actin filaments did not affect MCoTI-I uptake. As well, a significant amount of MCoTI-I accumulates in late endosomal and lysosomal compartments and MCoTI-I-containing vesicles continue to exhibit directed movements. These findings demonstrate internalization of MCoTI-I through endocytic pathways that are dominant in the cell type investigated, suggesting that this cyclotide has ready access to general endosomal/lysosomal pathways but could readily be re-targeted to specific receptors through addition of targeting ligands.

Letter of Submission

Dear Prof. Wim Hennink,

We have enclosed our manuscript entitled “Cellular Uptake of cyclotide MCoTI-I follows multiple endocytic pathways” by Janetter Contreras, Ahmed Elnagar, Sarah Hamm-Alvarez and myself, for your consideration to be published in the special issue on the 15th International Symposium on Recent Advances in Drug Delivery Systems that took place in Salt Lake City, USA, last February.

In this work we report the first analysis of intracellular uptake of MCoTI-I cyclotide in HeLa cells using using live cell imaging by confocal fluorescence microscopy. Cyclotides represent a novel new platform for drug development. Their stability, conferred by the cyclic cystine knot, their small size, their amenability to both chemical and biological synthesis and their flexible tolerance to sequence variation make them ideal for grafting of biologically-active therapeutic epitopes. As we show herein, they are also capable in the unmodified state of utilizing multiple cellular endocytic pathways for internalization. Their ease of access makes them readily accessible in their current state to endosomal/lysosomal compartments of virtually any cell. Without an apparent strong preference for an existing cellular pathway nor surface-expressed epitope in HeLa cells (nor in other studies with MCoTI-II in macrophages), they appear highly amenable to retargeting to exploit a particular target cell’s dominant internalization pathway and/or unique surface receptor repertoire, along with the targeted introduction of biologically-active therapeutic motifs.

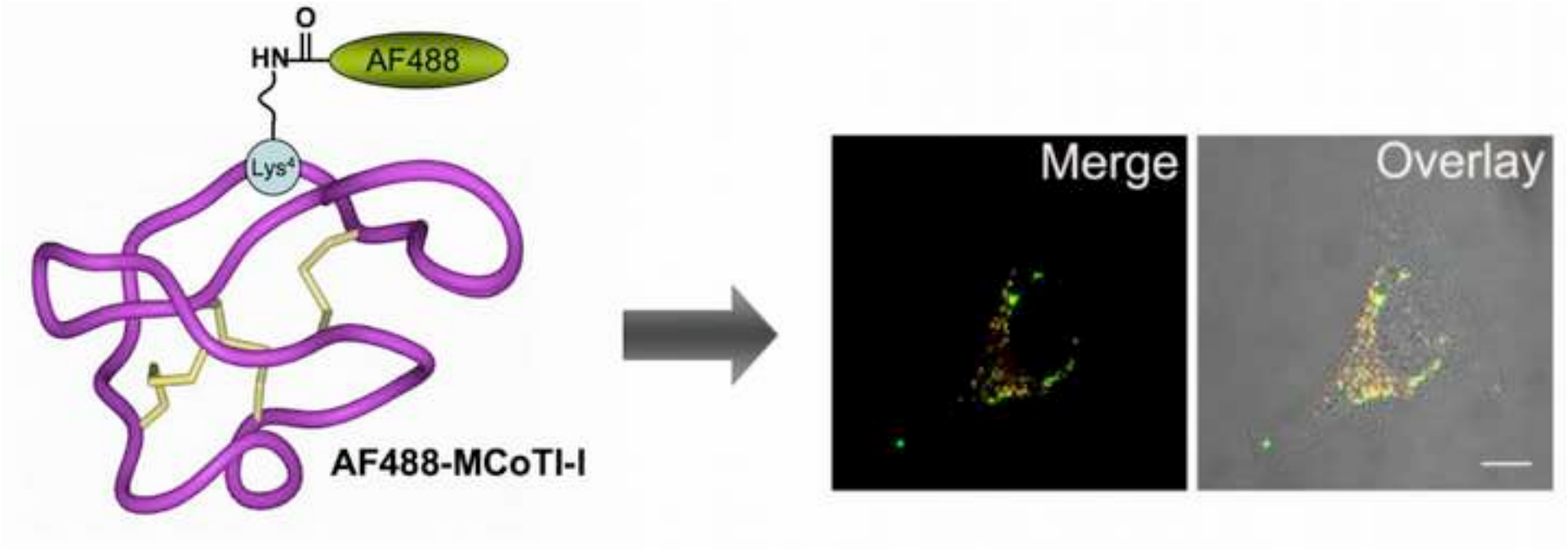
I am looking forward for your positive response.

All the graphics contained in our manuscript are in color. Please do not hesitate to contact me if you have any questions. I have included a short list of knowledgeable referees, which you may wish to contact.

Sincerely yours,

Julio A. Camarero, PhD
Associate Professor
Department of Pharmacology and Pharmaceutical Sciences
University of Southern California
Los Angeles, CA90033, USA
e-mail: jcamarer@usc.edu

*Graphical Abstracts



Cellular Uptake of cyclotide MCoTI-I follows multiple endocytic pathways

Janette Contreras[#], Ahmed Y. O. Elnagar[#], Sarah Hamm-Alvarez and Julio A. Camarero*

Department of Pharmacology and Pharmaceutical Sciences, School of Pharmacy, University of Southern California, Los Angeles, CA 90033, USA

[#]**Authors contributed equally to this work**

***Address all correspondence to:**

Julio A. Camarero, Ph.D.
Associate Professor
Department of Pharmacology and Pharmaceutical Sciences
University of Southern California
1985 Zonal Avenue, PSC 616
Los Angeles, CA 90033
E-mail: jcamare@usc.edu

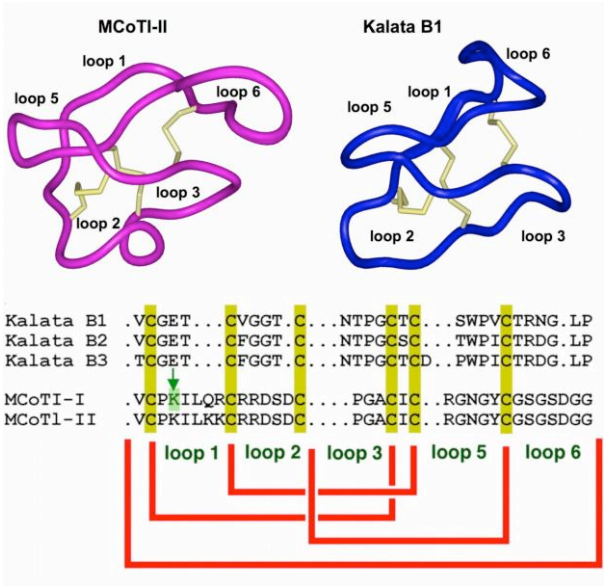
Abstract
Cyclotides are plant-derived proteins that naturally exhibit various biological activities and whose unique cyclic structure makes them remarkably stable and resistant to denaturation or degradation. These attributes, among others, make them ideally suited for use as drug development tools. This study investigated the cellular uptake of cyclotide, MCoTI-I in live HeLa cells. Using real time confocal fluorescence microscopy imaging we show that MCoTI-I is readily internalized in live HeLa cells and that its endocytosis is temperature-dependent. Endocytosis of MCoTI-I in HeLa cells is achieved primarily through fluid-phase endocytosis, as evidenced by its significant colocalization with 10K-dextran, but also through other pathways as well, as evidenced by its colocalization with markers for cholesterol-dependent and clathrin-mediated endocytosis, cholera toxin B and EGF respectively. Uptake does not appear to occur via macropinocytosis as inhibition of this pathway by Latrunculin B-induced disassembly of actin filaments did not affect MCoTI-I uptake. As well, a significant amount of MCoTI-I accumulates in late endosomal and lysosomal compartments and MCoTI-I-containing vesicles continue to exhibit directed movements. These findings demonstrate internalization of MCoTI-I through endocytic pathways that are dominant in the cell type investigated, suggesting that this cyclotide has ready access to general endosomal/lysosomal pathways but could readily be re-targeted to specific receptors through addition of targeting ligands.

Abbreviations: Boc, tert-butyloxy carbonyl; CCK, cyclic cystine knot; CPPs, cell-penetrating peptides; CTX-B, cholera toxin B; DAST, diethylaminosulfur trifluoride; DCM, dichloromethane; 10K-Dex, 10,000MW-dextran; DIEA, di-isopropylethylamine; DMF, dimethyl formamide; EGF, epidermal growth factor; ES-MS, electrospray-mass spectrometry; Fmoc, 9-fluorenyloxy carbonyl; HBTU, 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; HPLC, high performance liquid chromatography; Lat B, latrunculin B; MCoTI, *Momordica cochinchinensis* trypsin inhibitor; NMP, N-methyl-pyrrolidone; NMR, nuclear magnetic resonance; RFP-Lamp1, Red Fluorescent Protein-lysosomal associated protein 1; RP-HPLC, reverse phase-high performance liquid chromatography; TFA, trifluoroacetic acid; TIS, tri-isopropylsilane; TR, Texas Red; Trt, trityl.

Introduction

Cyclotides are fascinating micro-proteins ranging from 28 to 37 amino acid residues that are naturally expressed in plants and exhibit various biological activities such as anti-microbial, insecticidal, cytotoxic, antiviral (against HIV), and protease inhibitory activity, as well as exert hormone-like effects [1-4]. They share a unique head-to-tail circular knotted topology of three disulfide bridges, with one disulfide penetrating through a macrocycle formed by the two other disulfides and inter-connecting peptide backbones, forming what is called a cystine knot topology (Fig. 1). This cyclic cystine knot (CCK) framework gives the cyclotides exceptional resistance to thermal and chemical denaturation, and enzymatic degradation [4, 5]. In fact, the use of cyclotide-containing plants in indigenous medicine first highlighted the fact that the peptides are resistant to boiling and are apparently orally bioavailable [6].

Cyclotides have been isolated so far from plants in the *Rubiaceae*, *Violaceae*, *Cucurbitacea* [4, 7] and most recently *Fabaceae* families [8]. Around 160 different cyclotides sequences have been reported in the literature [9, 10], although it has been estimated that ≈ 50,000 cyclotides might exist [11, 12]. Despite the sequence diversity all cyclotides share the same CCK motif (Fig.1). Hence, these micro-proteins can be considered as natural combinatorial peptide libraries structurally constrained by the cystine-knot scaffold [2] and head-to-tail cyclization but in which hypermutation of essentially all residues is permitted with the exception of the strictly conserved cysteines that comprise the knot.



Cyclotides are ribosomally produced in plants from precursors that comprise between one and three cyclotide domains. However, the mechanism of excision of the cyclotide domains and ligation of the free N- and C-termini to produce the circular peptides has not been completely elucidated yet. It is suspected, however, that specific aspariginyl endopeptidases are involved in the proteolytic processing and cyclization of the precursor proteins [13-15]. Cyclotides can be also produced chemically using solid-phase peptide synthesis in combination with native chemical ligation [16-19] or recombinantly in bacteria by using a modified protein splicing units or inteins [20, 21]. The latter method can generate folded cyclotides either *in vivo* or *in vitro* using standard bacterial expression systems [20-22] and opens

Figure 1. Primary and tertiary structures of MCoTI and kalata cyclotides. The structures of MCoTI-II (pdb ID: 1IB9 [27] and kalata B1 (pdb ID: 1NB1 [82]) are shown. Conserved cysteine residues and disulfide bonds are shown in yellow. An arrow marks residue Lys^4 located at loop 1 in MCoTI-cyclotides. This residue in MCoTI-I was used for the site-specific conjugation of AlexaFluor488 N-hydroxysuccinimide ester (AF488-OSu) through an stable amide bond.

the possibility of producing large libraries of genetically encoded cyclotides which can be analyzed by high throughput cell-based screening for selection of specific sequences able to bind to particular biomolecular targets [21, 23].

Cyclotides have been classified into three main subfamilies. The Möbius and bracelet cyclotide subfamilies differ in the presence or absence of a *cis*-Pro residue, which introduces a twist in the circular backbone topology [24]. A third subfamily comprises the cyclic trypsin inhibitors MCoTI-I/II (Fig. 1), which have been recently isolated from the dormant seeds of *Momordica cochinchinensis*, a plant member of the *cucurbitaceae* family, and are powerful trypsin inhibitors (K_i ≈ 20 - 30 pM) [25]. These cyclotides do not share significant sequence homology with other cyclotides beyond the presence of the three-cystine bridges, but structural analysis by NMR has shown that they

adopt a similar backbone-cyclic cystine-knot topology [26, 27]. MCoTI cyclotides, however, show high sequence homology with related linear cystine-knot squash trypsin inhibitors [25], and therefore represent interesting molecular scaffolds for drug design [19, 28-30]. Indeed, acyclic squash inhibitors have been already used as scaffold for the incorporation of novel bioactive peptides to render de-novo engineered knottins with novel biological activities [31, 32].

All these features make cyclotides ideal drug development tools [19, 28-30]. They are remarkably stable due to the cyclic cystine knot [33]. They are relatively small, making them readily accessible to chemical synthesis [16]. They can also be encoded within standard cloning vectors, and expressed in cells [20-22], and are amenable to substantial sequence variation [34], which make them ideal substrates for molecular grafting of biological peptide epitopes [4] or amenable to molecular evolution strategies to enable generation and selection of compounds with optimal binding and inhibitory characteristics [22, 34]. Even more importantly, MCoTI-cyclotides have been shown recently to be able to enter human macrophages and breast cancer cell lines [35]. Internalization into macrophages was shown to be mediated mainly through macropinocytosis, a form of endocytosis that is actin-mediated and results in formation of large vesicles termed macropinosomes [36, 37]. It should be noted, however, that in this study the visualization of MCoTI-II uptake was done in fixed and not in live cells. Analysis of live cells provides the ability to visualize events in real time without the possible complications of fixation artifacts that have confounded interpretations of the uptake of Tat and other related peptides for instance [38, 39]. As well, macropinocytosis is a dominant mechanism for endocytic uptake in macrophages [40-42], unlike other cells that are not specialized for large scale sampling of extracellular fluid and which use multiple alternative endocytic mechanisms. These mechanisms can include clathrin-mediated endocytosis, caveolar endocytosis, macropinocytosis, phagocytosis, flotillin-dependent endocytosis, as well as multiple other as yet under-characterized mechanisms [43, 44]. Intrigued by these results, we explored the cellular uptake of site-specific fluorescent-labeled MCoTI-cyclotides and studied the cellular uptake mechanisms in HeLa cells using live cell imaging by confocal fluorescence microscopy.

In this work we report for the first time the cellular uptake of MCoTI-cyclotides monitored by real time confocal fluorescence microscopy imaging in live HeLa cells. Our results clearly show that HeLa cells readily internalize fluorescently-labeled-MCoTI-I. We found that this process is temperature-dependent and can be reversibly inhibited at 4°C, which indicates an active mechanism of internalization. The internalized cyclotide also seems to colocalize in live cells with multiple endocytic markers including, to the greatest extent, the fluid-phase endocytic marker dextran (10 KDa dextran,10K-Dex). Internalized MCoTI-I was colocalized to a lesser extent with the cholesterol/lipid dependent endocytic marker cholera toxin B (CTX-B) and the clathrin-mediated endocytic marker, EGF. Internalized MCoTI-I was localized within a fairly rapid time course with late endosomal and lysosomal compartments which engaged in rapid and directed movements suggestive of cytoskeletal involvement. MCoTI-I uptake in HeLa cells was not impaired by Latrunculin B (Lat B), a well-known inhibitor of macropinocytosis. Altogether, these data seem to indicate that MCoTI-I cyclotide is capable of internalization in live cells through multiple endocytic pathways that may be dominant in the particular cell type under study. The lack of strong preference for MCoTI-I internalization via a specific cellular internalization pathway is of significant value since the lack of endogenous affinity for a particular pathway can enable the ready re-targeting by introduction of targeting peptides within the scaffold that may enable specific and targeted endocytic uptake to a particular target cell. At the same time, the ready uptake of MCoTI-I by multiple pathways suggests accessibility, in the untargeted form, to essentially all cells.

Materials and Methods

Analytical characterization of cyclotides: Analytical HPLC was performed on a HP1100 series instrument with 220 and 280 nm detection using a Vydac C18 column (5 micron, 4.6 x 150 mm) at a flow rate of 1 mL/min. Preparative and semi-preparative HPLC were performed on a Waters

Delta Prep system fitted with a Waters 2487 UV-visible detector using a Vydac C18 (15-20 µm, 10 x 250 mm) at a flow rate of 5 mL/min. All runs used linear gradients of 0.1% aqueous trifluoroacetic acid (TFA, solvent A) *vs.* 0.1% TFA, 90% acetonitrile in H_2O (solvent B). Ultraviolet-visible (UV-vis) spectroscopy was carried out on an Agilent 8453 diode array spectrophotometer. Electrospray mass spectrometry (ES-MS) analysis was routinely applied to all compounds and components of reaction mixtures. ES-MS was performed on an Applied Biosystems API 3000 triple quadrupole electrospray mass spectrometer using Analyst 1.4.2. Calculated masses were obtained using Analyst 1.4.2. All chemicals involved in synthesis or analysis were obtained from Aldrich (Milwaukee, WI) or Novabiochem (San Diego, CA) unless otherwise indicated.

Preparation of Fmoc-Tyr(tBu)-F. Fmoc-Tyr(tBu)-F was prepared using diethylaminosulfur trifluoride DAST [45] and quickly used afterwards. Briefly, to a stirred solution of Fmoc-Tyr(tBu)-OH (459.6 mg, 1 mmol) in 10 mL of dry dichloromethane (DCM), containing dry pyridine (800 µL, 1 mmol) and (1.1 mL, 1.2 mmol) of DAST was added dropwise at 25° C under nitrogen current. After 20 minutes, the mixture was washed with ice-cold water (3 x 20 mL). The organic layer was separated and dried over anhydrous $MgSO_4$. The solvent was removed under reduced pressure to give the corresponding Fmoc-amino acyl fluoride as white solid that was used immediately. Amino acid fluorides should be used immediately as they are extremely unstable and prone to hydrolysis.

Loading of 4-sulfamylbutyryl AM resin with Fmoc-Tyr(tBu)-F. Loading of the first residue was accomplished using Fmoc-Tyr(tBu)-F according to standard protocols [46]. Briefly, 4-Sulfamylbutyryl AM resin (420mg, 0.33 mmol) (Novabiochem) was swollen for 20 minutes with dry DCM and then drained. A solution of Fmoc-Tyr(tBu)-□□1 □□□□□ □□□□□ □ □□□□□□F4□□ □ □ □□□□□ □□□□□□□ di-isopropylethylamine (DIEA) (180 µL, 1 mmol) was added to the drained resin and reacted at 25° C for 1 h. The resin was washed with dry DCM (5 x 5 mL), dried and kept at m20°C until use.

Chemical synthesis of MCoTI-I. Solid-phase synthesis was carried out on an automatic peptide synthesizer ABI433A (Applied Biosystems) using the Fast-Fmoc chemistry with 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) activation protocol at 0.1 mmole scale on a Fmoc-Tyr(tBu)-sulfamylbutyryl AM resin. Side-chain protection was employed as previously described for the synthesis of peptide α-thiesters by the Fmoc-protocol [47], except for the N-terminal Cys residue, which was introduced as Boc-Cys(Trt)-OH. After chain assembly, the alkylation, thiolytic cleavage and deprotection were performed as previously described [48, 49]. □F□□4□1 □□□□□ □□□□□□F□(□□(□□□□□□□(□□□□F□ □□≈ □F□□□F (□□□4□□(□□□(≈ □□(□□ □ □≈ (□□□□ $□_2$CN (174 µL, 2.4 mmol; previously filtered through basic silica) and DIEA (82 µL, 0.46 mmol) in N-methylpyrrolidone (NMP) (2.2 mL) for 12 h. The resin was then washed with NMP (3 x 5 mL) and DCM (3 x 5 mL). The alkylated peptide resin was cleaved with $HSCH_2CH_2CO_2Et$ (200 µL, 1.8 mmol) in the presence of a catalytic amount of sodium thiophenolate (NaSPh, 3 mg, 22 µmol) in dimethylformamide (DMF):DCM (3:4 v/v, 1.4 mL) for 24 h. The resin was then dried at reduced pressure. The side-chain protecting groups were removed by treating the dried resin with trifluoroacetic acid (TFA):H_2O:tri-isopropylsilane (TIS) (95:3:2 v/v, 5 mL) for 3-4 h at room temperature. The resin was filtered and the linear peptide thioester was precipitated in cold Et_2O. The crude material was dissolved in the minimal amount of H_2O:MeCN (4:1) containing 0.1% TFA and characterized by HPLC and ES-MS as the desired MCoTI-I linear precursor α-thioester [Expected mass (average isotopic composition) = 3608.2 Da; measured = 3608.8 ± 0.3 Da]. Cyclization and folding was accomplished by flash dilution of the MCoTI-I linear α-thioester TFA crude to a final concentration □□1 □□6□ □□(□□F□ □□4□□□□ □□2 mM reduced glutathione (GSH), 50 sodium phosphate buffer at pH 7.5 for 18 h. Folded MCoTI-I was purified by semi-preparative HPLC using a linear gradient of 10-35% solvent B over 30 min. Pure MCoTI-I was characterized by HPLC and ES-MS [Expected mass (average isotopic composition) = 3480.9 Da; measured = 3481.0 ± 0.4 Da].

Recombinant Expression of MCoTI-I. Bacterial expression and purification of MCoTI-I was carried out reviously described [22].

Chemical labeling of MCoTI with AlexaFluor488 succinimide ester (AF488-NHS). MCoTI-I was site-specifically labeled through the ε-amino of residue Lys^4 (Fig.1). MCoTI-I only has one Lys residue in its sequence (Fig. 1). Briefly, MCoTI-I (5 mg, 1.4 µmol) was conjugated with two-fold molar excess of AF488-MHS in 0.2 M sodium phosphate buffer (2.5 mL) at pH 7.5 for 2 h. The reaction was quenched with 6 mM NH_2-OH solution at pH 4. AF488-labeled MCoTI-I was purified by semi-preparative HPLC using a linear gradient of 15-35% solvent B over 30 min. Pure labeled MCoTI-I was characterized by HPLC and ES-MS [Expected mass (average isotopic composition) = 3997.9 Da; measured = 3997.4 ± 0.3 Da].

Purification of synthetic MCoTI-I using trypsin-Sepharose beads. Preparation of trypsin-Sepharose beads was done as previously described [21-23]. Pull down experiments with synthetic MCoTI-I were performed as follows: Synthetic MCoTI-I cyclization/folding crude reactions were typically incubated with 0.2 mL of trypsin-Sepharose for one hour at room temperature with gentle rocking, and centrifuged at 3000 rpm for 1 min. The beads were washed with 50 volumes of PBS containing 0.1% Triton X-100, then rinsed with 50 volumes of PBS, and drained of excess PBS. Bound MCoTI-I was eluted with 0.4 mL of 8 M GdmCl and fractions were analyzed by RP-HPLC and ES-MS.

Endocytosis Experiments. For studies of endocytic uptake mechanisms, methyl-F-cyclodextrin (MBCD) was purchased from Sigma-Aldrich. Latrunculin B (Lat B) was purchased from Calbiochem . Texas Red-EGF (TR-□□ □□□□β□□□□□□□E □□ □□□□N-99 (LysoRed), AF594 cholera toxin B (AF594 CTX-B), Texas Red 10,000 MW dextran (TR-□□□□□□□□□□□□□□□□□□E Lysosomes-RFP (RFP-Lamp1), rhodamine-phalloidin, and DAPI were all purchased from Invitrogen (Carlsbad, CA).

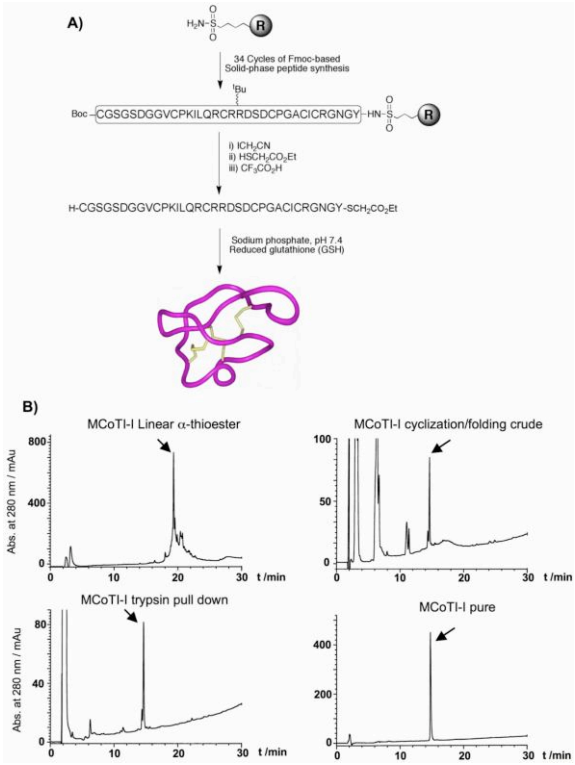
Cell Culture. HeLa cells were obtained from the American Type Culture Collection (ATCC) and were cultured in a humidified incubator at 37°C in 95% air/5% CO_2 in phenol red-□□□□□□□□□□G□ modified essential medium (DMEM)(4.5 g/L glucose with 10% FBS, 1% glutamine, and 1% non-essential amino acids) and split with trypsin/EDTA as recommended by the manufacturer.

Confocal Fluorescence Microscopy. For MCoTI-I uptake studies, HeLa cells were seeded on 35 mm glass-bottom culture dishes (MatTek, Ashland, MA) at a density of 8.5 X 10^4 cells/dish. On day 2 of culture, the cells were rinsed with PBS and the media replaced with incubation buffer (phenol red-free, serum-free DMEM with 1% P/S and 20 mM HEPES) prior to addition of AF488-MCoTI-I (25 µM) and incubation at 37°C for 1 hr. Following this time, excess MCoTI-I was rinsed off with a gentle PBS wash and the media replaced prior to imaging. Intracellular distribution was analyzed at 1 hr and again at regular intervals for up to 10 hours. For assessment of distribution and colocalization of AF488-MCoTI-I and LysoTracker™ Red, RFP-Lamp1, AF594-Cholera toxin B, TR-10K Dex, or TR-EGF in live cells, we utilized a Zeiss LSM 510 Meta NLO imaging system equipped with Argon and HeNe lasers and mounted on a vibration-free table for confocal fluorescence microscopy. For analysis of the effects of Lat B pre-treatment, 2 µM Lat B was added for 30 min at 37°C prior to addition of AF488-MCoTI-I. For colocalization studies, LysoTracker™ Red (50 nM), AF594-Cholera toxin B (10 µg/mL), TR-10K Dex (1 mg/mL), or TR-EGF (400 ng/mL) was added to cells simultaneously with MCoTI-I prior to incubation at 37°C. Analysis of the extent of colocalization was done at 1 hr of uptake. For colocalization with RFP-Lamp1, cells were treated with RFP-Lamp1-expressing BacMam (2 x 10^7 particles/plate) on the previous day. For temperature-dependent uptake studies, cells were cooled on ice for 30 min prior to the addition of AF488-MCoTI-I in incubation buffer. After incubation at 4°C for 30 min, the cells were imaged and subsequently incubated at 37°C for 1 hour before imaging again. For fixation and visualization of

actin filaments following treatment with or without 2 µM Lat B, the cells were fixed with 4% paraformaldehyde prior to the addition of rhodamine-phalloidin and DAPI. For analysis of fluorescent pixel colocalization, cells from at least 3 different experiments were analyzed individually. Using the Zeiss LSM 510 software colocalization tool, regions of interest (ROI) were selected and marked with an overlay to encompass all pixels, following the Zeiss manual protocol. The threshold was automatically set from these ROIs. For time-lapse imaging, cells were incubated with 25 µM AF488-MCoTI-I for 1 h at 37°C. Following this time, excess MCoTI-I was rinsed off with a gentle PBS wash and the media replaced prior to imaging. The time series image capture was set to a 2.5 second delay between scans.

Results and Discussion

In order to study the cellular uptake of MCoTI-cyclotides, we decided to use MCoTI-I. MCoTI-I contains only one Lys residue located in loop 1 versus MCoTI-II, which contains three Lys residues in the same loop (Fig. 1). The presence of only one Lys residue facilitates the site-specific introduction of a unique fluorophore on the sequence thus minimizing any affect that the introduction of this group may have on the cellular uptake properties of the cyclotide.



Folded MCoTI-I cyclotide was produced either by recombinant or synthetic methods. In both cases the backbone cyclization was performed by an intramolecular native chemical ligation (NCL) [50-53] using the native Cys located to the beginning of loop 6 to facilitate the cyclization. This ligation site has been shown to give very good cyclization yields [21, 22]. Intramolecular NCL requires the presence of an N-terminal Cys residue and C-terminal α-thioester group in the same linear precursor [52, 54]. In the biosynthetic approach, the MCoTI-I linear precursor was fused in frame at their C- and N-terminus to a modified Mxe Gyrase A intein and a Met residue, respectively and expressed in *Escherichia coli* [23]. This allows the generation of the required C-terminal thioester and N-terminal Cys residue after *in vivo* processing by endogenous Met aminopeptidase (MAP) [20, 55]. Cyclization and folding can be accomplished very efficiently *in vitro* by incubating the MCoTI-I intein fusion construct in sodium phosphate buffer at pH 7.4 in the presence of reduced glutathione (GSH). Biosynthetic MCoTI-cyclotides generated this way have been shown to adopt a native folded structure by NMR and trypsin inhibitory assays [20, 22, 33].

Figure 2. Chemical synthesis of MCoTI-I (A) Synthetic scheme used for the chemical synthesis of cyclotide MCoTI-I by Fmoc-based solid-phase peptide synthesis (B) Analytical reverse-phase HPLC traces of MCoTI-I linear precursor α-thioester, cyclization/folding crude and purified MCoTI-I by either affinity chromatograpgy using trypsin-immobilized Sepharose beads or semipreparative reverse-phase HPLC. HPLC analysis was performed in all the cases using a linear gradient of 0% to 70% buffer B over 30 min. Detection was carried out at 220 nm. An arrow indicated the desired product in each case.

Natively folded MCoTI-II has been already successfully produced using Fmoc-based solid-phase peptide synthesis [18, 19]. Encouraged by these results we also explored the production of MCoTI-I by chemical synthesis (Fig. 2). For this purpose the MCoTI-I linear precursor α-thioester was assembled by Fmoc-based solid-phase peptide synthesis on a sulfonamide resin [48, 49] (Fig. 2A).

Activation of the sulfonamide linker with iodoacetonitrile followed by cleavage with ethyl mercaptoacetate and acidolytic deprotection with TFA provided the fully protected linear peptide α-thioester (Fig. 2B). The synthetic linear precursor thioester was then efficiently cyclized and folded in one-pot reaction using sodium phosphate buffer at pH 7.5 in the presence of 2 mM GSH. The reaction was complete in 18 h and the folded product was purified by reverse-phase HPLC and characterized by ES-MS. The expected mass for folded MCoTI-I was in agreement with a folded structure (Expected mass = 3480.9 Da; measured = 3481.0 ± 0.4 Da). Synthetic folded MCoTI-II was also shown to co-elute by HPLC with recombinant natively folded MCoTI-I (data not shown). The biological activity of synthetic MCoTI-I was assayed by using a trypsin pull-down experiment [22, 23]. As shown in Figure 2B, synthetic folded MCoTI-I was specifically captured from a cyclization/folding crude reaction by trypsin-immobilized Sepharose beads [21-23], thus indicating that was adopting a native cyclotide fold. Purified MCoTI-I was site-specifically labeled with AlexaFluor 488 (AF488) for live confocal imaging. The ε-amino group of Lys^4 residue located in loop 1 was conjugated to AF488-NHS in sodium phosphate buffer at pH 7.5 for 2 h (Fig. 3A). Under these conditions the main product of the reaction was mono-labeled AF488-MCoTI as characterized by HPLC and ES-MS (expected average mass = 3997.9 Da; measured = 3997.4 ± 0.3 Da) (Figs 3C). AF488-labeled MCoTI-I was then purified by reverse-phase HPLC to remove any trace of unreacted materials (Fig. 3B).

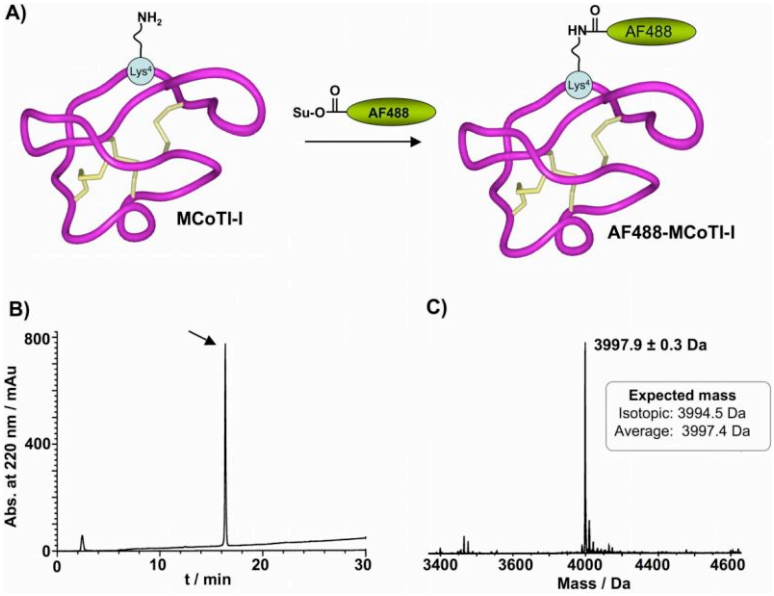
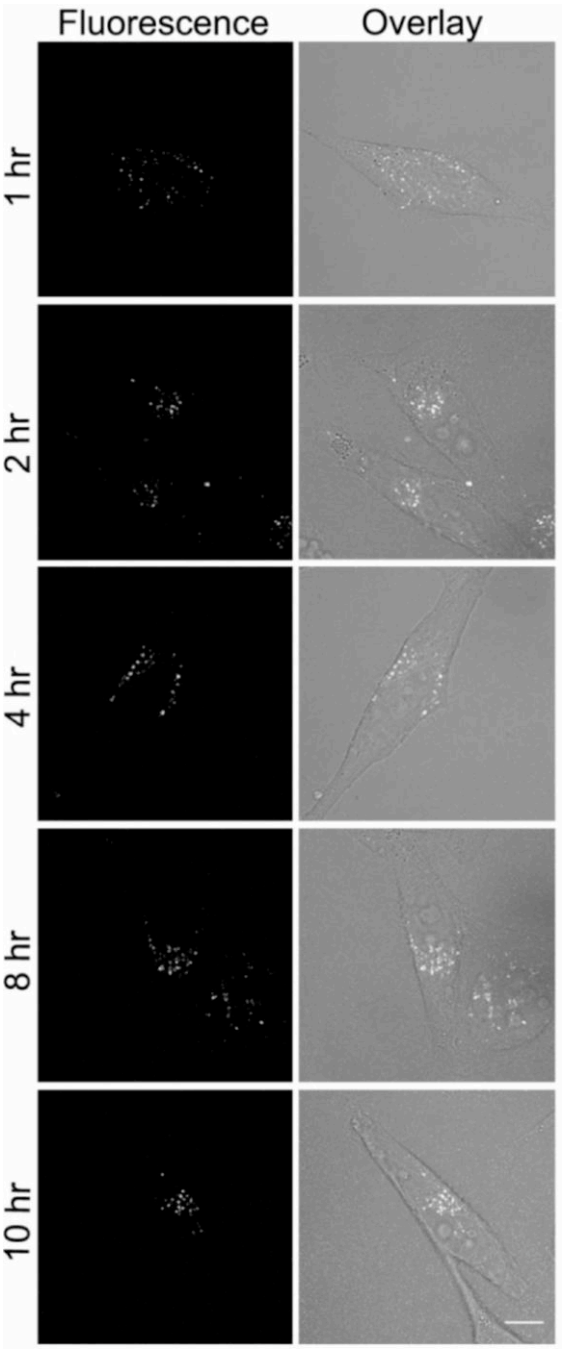


Figure 3. Site-specific labeling of MCoTI-I with AlexaFluor-488 N-hydroxysuccinimide ester (AF488-OSu). (A) Scheme depicting the bioconjugation process and localization of the fluorescent probe at residue Lys^4 in loop 1. (B) Analytical reverse-phase HPLC trace of pure AF488-MCoTI-I. HPLC analysis was performed using a linear gradient of 0% to 70% buffer B over 30 min. Detection was carried at 220 nm (C) ES-MS spectra of pure AF488-MCoTI-I.

In order to infer the correct conclusions regarding data obtained on the cellular uptake of native MCoTI-I when using modified cyclotides, like AF488-MCoTI-I for example, it is critical to be sure that they still adopt structures similar to that of the native form. MCoTI-cyclotides are extremely stable to chemical and thermal denaturation, and they have been shown to be able to withstand procedures like reverse-phase chromatography in the presence of organic solvents under acidic conditions without affecting their tertiary structure [16, 18-21, 33]. It is also unlikely that the acylation of the ε-amino group of Lys^4 in MCoTI-I may disrupt the tertiary structure of this cyclotide. Craik and co-workers have previously shown that biotinylation of the three Lys residues located in loop 1 in MCoTI-II, including Lys^4 (Fig. 1) does not disrupt the native cyclotide fold of this cyclotide as determined by ^{1}H-NMR [35]. We have also recently shown that mutation of residue Lys^4 by Ala does not seem to affect the ability of this mutant to adopt a native cyclotide fold, thus indicating that the presence of positive charge residue in this position is not critical for the tertiary structure of MCoTI-I [22]. Similar findings have been also found by Leatherbarrow and coworkers, where mutation of this residue by Phe or Val was still able to render MCoTI-cyclotides able to fold correctly and have inhibitory activity against chymotrypsin and human elastase, respectively [19]. Altogether these facts suggest that residue Lys^4 is not critical for adopting the native cyclotide fold or disturbing the tertiary structure of MCoTI-cyclotides.



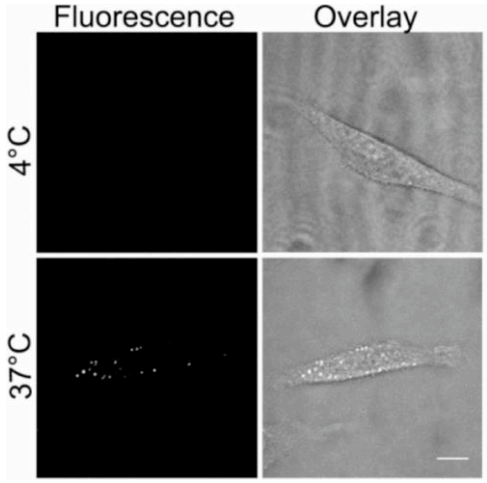
To study the cellular uptake of AF488-MCoTI-I we used HeLa cells. The internalization studies were all carried out with 25 µM AF488-MCoTI-I. This concentration provided a good signal/noise ratio for live cell confocal fluorescence microscopy studies and did not show any cytotoxic effect on HeLa cells. This is in agreement with the cellular tolerance of wild-type and biotinylated MCoTI-II reported for other types of human cell lines [35]. First, we analyzed the time course of changes in cellular distribution following uptake of 25 µM AF488-MCoTI-I by incubating with the cyclotide for 1 hr and then analyzing its distribution after 1, 2, 4, 8 and 10 h. As shown in Figure 4, the internalized cyclotide was clearly visible within perinuclear punctate spots inside the cells after 1 h incubation. Observation of cells pulsed with AF488-MCoTI-I for one hour and then incubated for longer periods of time in the absence of cyclotide did not show any evidence for decreased intracellular fluorescence, while the largely perinuclear distribution of internalized MCoTI-I appeared comparable at all time points. Similar results have been also been recently reported on the internalization of biotinylated MCoTI-II in macrophage and breast cancer cell lines [35], these studies however, used fixed cells to visualize the internalized cyclotide.

In order to study the mechanism of internalization of AF488-MCoTI-I in live HeLa cells, we first explored the effect of temperature on the uptake process. Active and energy-dependent endocytic mechanisms of internalization are inhibited at 4° C [56]. The internalization of AF488-MCoTI-I was totally inhibited after a 1 h incubation at 4°C (Fig. 5). This inhibition was completely reversible and when the same cells were incubated again at 37°C for 1 h, the punctate intracellular fluorescence labeling pattern was restored. This result confirmed that the uptake of AF488-MCoTI-I in HeLa cells follows a

Figure 4. MCoTI-I distribution in HeLa cells. HeLa cells were incubated with 25 µM MCoTI-I for 1 hour, cyclotide was removed with gentle rinsing in PBS and then the cells were monitored for distribution of intracellular fluorescence at intervals from 1- 10 hours using confocal fluorescence microscopy. Bar = 10µm.

temperature dependent active endocytic internalization pathway. It should be noted that no significant surface binding was detected at 4°C, suggesting that MCoTI-I does not bind a surface receptor, even nonspecifically. This is in agreement with studies on the MCoTI-II in fixed cells so both the MCoTI-I and MCoTI-II appear to lack specific affinity for proteins or lipids in cell membranes, unlike the kalata B1 cyclotide which shows membrane affinity [35]. This lack of endogenous affinity for a specific surface receptor or membrane constituent makes MCoTI-I ideal for engineering using more specific, receptor-directed, peptide-based, internalization motifs, within

the scaffold, that might enable members of this family to have targeting enhanced to a specific cell type.



Next, we investigated the internalization pathway used by labeled-MCoTI-I to enter HeLa cells. There are several known and well-characterized mechanisms of endocytosis [57]. It is also now well established that almost all cell-penetrating peptides (CPPs) use a combination of different endocytic pathways rather than a single endocytic mechanism [57]. A recent study showed that several CPPs (including Antennapedia/penetratin, nona-Arg and Tat peptides) can be internalized into cells by multiple endocytic pathways including macropinocytosis, clathrin-mediated endocytosis, and caveolae/lipid raft mediated endocytosis [58]. To investigate if that was the case with the internalization of AF488-MCoTI-I in HeLa

Figure 5. Endocytosis of MCoTI-I is temperature-dependent. HeLa cells were incubated with 25 µM MCoTI-I for 1 hr at 4°C. After removal of the MCoTI-I-containing media, and a gentle PBS wash, the cells were imaged. Following imaging, the MCoTI-I-containing media was replaced and the cells incubated at 37°C for 1 hr and imaged again. Bar = 10 µm.

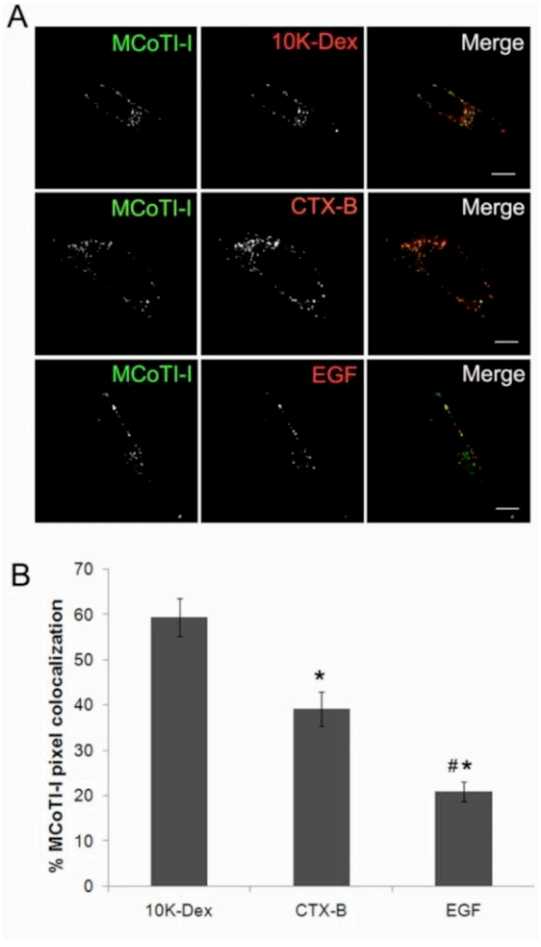
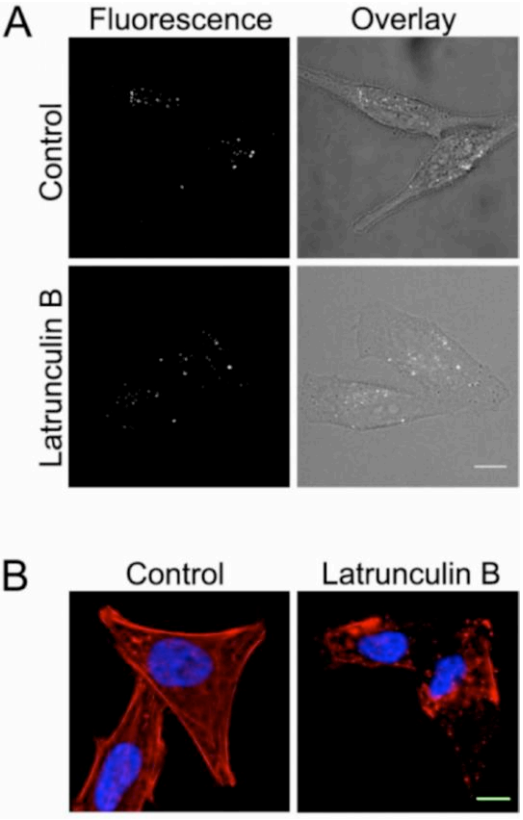


Figure 6. Colocalization of MCoTI-I with markers of endocytosis. (A) HeLa cells were incubated with 25 µM MCoTI-I and either 1 mg/ml 10K-dextran (10K-Dex), 10 µg/ml cholera toxin B (CTX-B), or 400 ng/ml epidermal growth factor (EGF) for 1 hour at 37°C as described in Materials and Methods and then imaged. Bar = 10 µm. (B) Quantification of pixel colocalization was done using the Zeiss LSM software for image analysis and measures the % of total fluorescent AF488 MCoTI-I pixels in the ROI relative to red pixels associated with different endocytic markers. (n = 13 cells for 10K-Dex, n = 11 cells for CTX-B and n = 10 cells for EGF, with cells assessed across 3 different experiments, * $p \leq 0.05$ relative to 10K-Dex, # $p \leq 0.05$ relative to CTX-B).

cells, we decided to look at its colocalization with various endocytic markers (Fig. 6). 10K-Dex has previously been used as a marker of fluid-phase endocytosis [36, 59-61]. CTX-B has been used as a marker for various lipid-dependent endocytic pathways [44, 62], while EGF has traditionally been a marker of clathrin-mediated endocytosis [63-65]. As shown in Figure 6, colocalization studies showed that after 1 h, AF488-MCoTI-I fluorescence was significantly colocalized with the fluorescence associated with 10K-Dex (59 ± 4% of total cyclotide fluorescence pixels were colocalized with 10K-Dex fluorescent pixels). Less colocalization was observed with fluorescent CTX-B (39 ± 4 %) and fluorescent EGF (21 ± 2 %). This data seems to suggest that AF488-MCoTI-I is primarily entering cells through fluid-phase endocytosis. The observed traces of colocalization with CTX-B and EGF also suggest that AF488-

MCoTI-I could be using alternative or additional endocytic pathways. The colocalization results could also be attributed, however, to the merging of endosomal uptake vesicles generated by different pathways at the level of an early endosome. To address whether the major uptake and colocalization of AF488-MCoTI-I with 10K-Dex was due to cointernalization by macropinocytosis, we explored the inhibition of AF488-MCoTI-I uptake by Lat B, a potent inhibitor of actin polymerization, which is an essential element of macropinocytlsis [66-69]. As shown in Figure 7, Lat B did not significantly inhibit uptake of AF488-MCoTI-I (Fig. 7A) nor of 10K-Dex (data not shown). Treatment of HeLa cells with this agent resulted in a total disruption of the actin filament network (Fig. 7B). These data suggest that macropinocytosis is not responsible for uptake of either 10K-Dex nor AF488-MCoTI-I in HeLa cells.



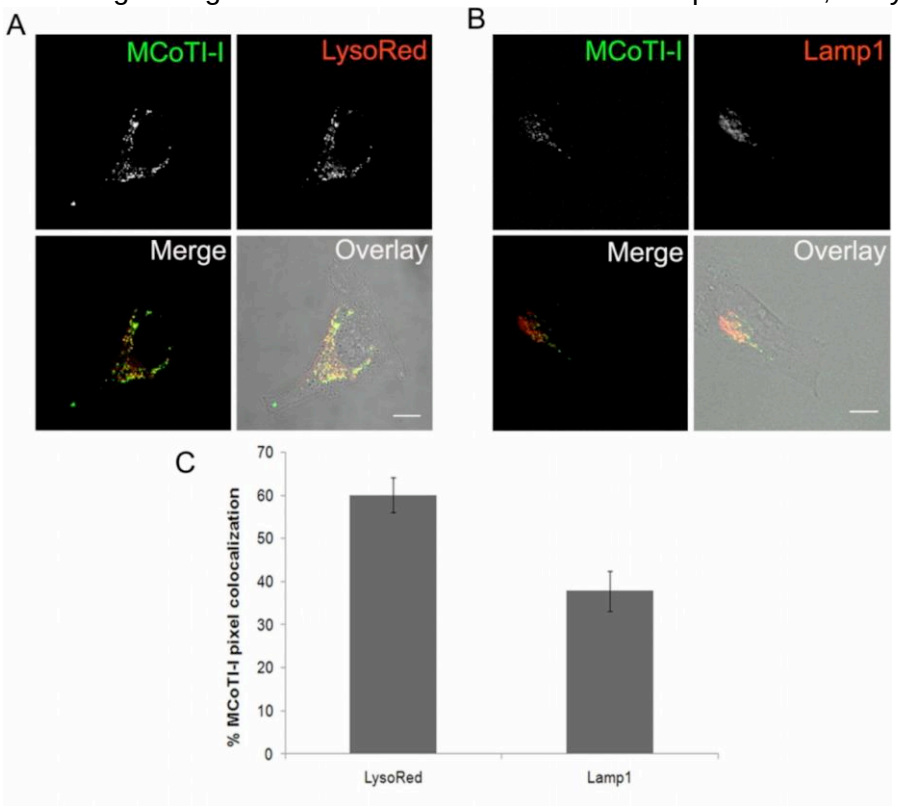
As an extension of these inhibition studies, cells were also treated with MBCD, a well-established cholesterol-depleting agent employed for studying the involvement of lipid rafts/caveolae in endocytosis [70, 71]. Preliminary studies with MBCD suggested no significant inhibition of AF488-MCoTI-I (data not shown). Since the extent of total colocalization of AF488-MCoTI-I with CTX-B was less than 40%, it is unsurprising that no marked effect was seen by live cell microscopy. Taken together, these results seem to suggest that the uptake of AF488-MCoTI-I in HeLa cells is following multiple endocytic pathways, which is in agreement with what has been recently reported for different CPPs [58].

Next we explored the fate of the endocytic vesicles containing labeled MCoTI-I. There are at least two pathways that involve the cellular trafficking of endosomal vesicles. The degradative pathway includes routing of internalized materials from early endosomes via late endosomes to lysosomes where degradation of internalized materials occurs within the cells. On the other hand, recycling endosomes sort material internalized into early endosomes and are responsible for effluxing internalized material back to the cellular membrane [72]. If labeled-MCoTI-I was localized in recycling endosomes, it would be expected that its concentration in the cell would

Figure 7. Disruption of actin does not inhibit MCoTI-I uptake. (A) HeLa cells were untreated (control) or treated with Lat B (2 µM) for 30 min at 37°C prior to addition of 25 µM MCoTI-I. Following uptake for 1 hr at 37°C, the cells were imaged using confocal fluorescence microscopy. Bar = 10 µm. (B) HeLa cells without treatment (control) or treated with 2 µM Lat B for 30 min at 37°C were fixed and labeled with rhodamine—phalloidin to label actin (red) and DAPI to label nuclei (blue). Bar = 10 µm.

decrease and/or accumulate on the membrane over time, which was not the case in the time course experiment following the cellular fate of internalized cyclotide (Fig. 4). To explore the potential localization of labeled-MCoTI-I in lysosomes we first used LysoTracker Red (LysoRed). This pH sensitive fluorescent probe is utilized for identifying acidic organelles, such as lysosomes and late endosomes, in live cells. As shown in Figure 8A, significant colocalization (60 ± 4.0% as determined by pixel colocalization analysis) of LysoRed and AF488-MCoTI-I was observed after treating the cells for 1h with both agents. As an extension of these experiments, we also investigated the colocalization of labeled-MCoTI-I and lysosomal-associated membrane protein 1 (Lamp1), an established mature lysosomal marker [73, 74]. For this experiment, live HeLa cells were first infected with a Red Fluorescent Protein (RFP)-Lamp1-expressing BacMam virus. The

next day the cells were incubated with AF488-MCoTI-I for 1 h and imaged. As shown in Figure 8B, colocalization was also seen for AF488-MCoTI-I and RFP-Lamp1 (38 ± 5%, as determined by pixel colocalization analysis), suggesting that even after 1 h, significant MCoTI-I has already reached the lysosomal compartments. Our data suggest that after 1 h, a significant amount of MCoTI-I (≈40%) has trafficked through the endosomal pathway to the lysosomes and that ≈20% is already localized in late endosomes or other types of acidic organelles. It has previously been reported that the perinuclear steady-state distribution of lysosomes is a balance between movement on microtubules and actin filaments [75-77]. Likewise, movement from early endosomal compartments to late endosomes to lysosomes has also been shown to rely on the microtubule network [78, 79]. As an extension of these experiments, and to investigate whether MCoTI-I-containing vesicles were actively trafficking inside the cell, we captured time-lapse video of cells after incubation with MCoTI-I for 1 h. Indeed, the time-lapse capture showed active movements of MCoTI-I-containing vesicles (Fig. 9). Directed short- and long-range movements could be seen, characteristic of movement on cytoskeletal filaments. These results suggest that while a large portion of MCoTI-I has reached lysosomal compartments by 1 h, and some of the movements seen may be attributed to the steady-state distribution of lysosomes, the remaining cyclotide may still be trafficking through the cell from other membrane compartments, likely within late endosomes.



Craik and co-workers have recently reported the uptake of biotinylated-MCoTI-II by human macrophages and breast cancer cell lines [35]. This work concluded that the uptake of MCoTI-II in macrophages is mediated by macropinocytosis and that the cyclotide accumulates in macropinosomes without trafficking to the lysosome. MCoTI-II shares high homology with MCoTI-I (≈97% homology, see Fig. 1) and similar biological activity. Despite their similarities, the differences in the cellular uptake and trafficking of MCoTI-cyclotides by macrophages versus HeLa cells could be attributed to the cellular

Figure 8. MCoTI-I is colocalized with lysosomal compartments. (A) Untreated or BacMam-RFP-Lamp1 treated HeLa cells were incubated with 25 µM MCoTI-I and LysoTracker Red, or MCoTI-I alone, for 1 hr at 37°C as described in Materials and Methods and then imaged. Bar = 10 µm. (B) Quantification of % of total fluorescent AF488-MCoTI-I pixel colocalization with fluorescent pixels associated with both markers was done using the Zeiss LSM software for image analysis. (*n* = 14 cells for LysoTracker Red and *n* = 11 cells for Lamp1 with cells selected from 3 separate experiments).

differences in endocytic preferences for these two very different cell types. Macrophage cells are specialized in large scale sampling of extracellular fluid using macropiniocytosis as the dominant endocytic pathway. Meanwhile, other types of cells may use multiple endocytic pathways as has been recently shown for the uptake of different CPPs in HeLa cells [58].

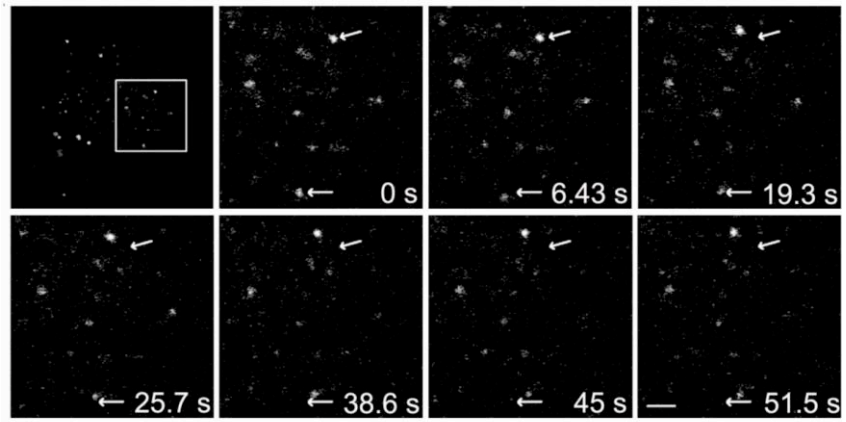


Figure 9. MCoTI-I-containing vesicles are in motion. HeLa cells were incubated with 25 μM MCoTI-I for 1 hr at 37°C and then imaged using time-lapse microscopy as described in Materials and Methods. Arrows indicate position of the moving vesicle at 0 min while displacement of the fluorescent vesicle relative to the arrow shows the extent of movement over time. Bar = 2 μm.

At this point we cannot be certain if some labeled MCoTI-I is able to escape from endosomal/lysosomal compartments into the cytosol. The ability to track the release of fluorescent-labeled molecules from cellular vesicles is limited using live cell imaging of fluorescence signal primarily due to the large dilution effect if the molecule is able to escape the highly confined volume of the vesicle into the larger cytosolic volume. One way to demonstrate the release of peptide into the cytosol, however, would be by using labels with better detection sensitivity or incorporating a biological activity that can be measured in the cellular cytosol. For example, the presence of Tyr residues in both MCoTI-cyclotides should facilitate the incorporation of radioactive iodine into the phenolic ring of Tyr with minimal disruption of the native structure of the cyclotide. The incorporation or grafting of biological peptides into the MCoTI scaffold could also provide proof of endosomal/lysosomal escape if such biological activity could be measured only in the cytosol. This approach has been already used to demonstrate endosomal escape of CPPs such as the TAT peptide [80, 81]. The retention of fluorescence signal in the perinuclear, lysosomal compartments for a period of up to 10 hrs suggests that most of the cyclotide remains within these compartments however; given the flexibility of the cyclotide backbone to accommodate multiple peptide sequences, subsequent studies may explore the ability of targeting and endosomolytic sequences for concomitant targeted entry and endosomal/lysosomal escape into cytosol.

Conclusion

This study reports on the first analysis of intracellular uptake of MCoTI-I cyclotide using live cell imaging by confocal fluorescence microscopy. Cyclotides represent a novel new platform for drug development. Their stability, conferred by the cyclic cystine knot, their small size, their amenability to both chemical and biological synthesis and their flexible tolerance to sequence variation make them ideal for grafting of biologically-active therapeutic epitopes. As we show herein, they are also capable in the unmodified state of utilizing multiple cellular endocytic pathways for internalization. Their ease of access makes them readily accessible in their current state to endosomal/lysosomal compartments of virtually any cell. Without an apparent strong preference for an existing cellular pathway nor surface-expressed epitope in HeLa cells (nor in other studies with MCoTI-II in macrophages [35]), they appear highly amenable to retargeting to exploit a particular target cell's dominant internalization pathway and/or unique surface receptor repertoire, along with the targeted introduction of biologically-active therapeutic motifs.

Acknowledgements:

The authors acknowledge the support of NIH grants EY017293 (to SHA), GM090323 (to JAC) and the support of a Ruth Kirchstein NIH minority predoctoral fellowship (F31EY018807) to JC.

Bibliography

[1] D.J. Craik, S. Simonsen, N.L. Daly, The cyclotides: novel macrocyclic peptides as scaffolds in drug design. Curr. Opin. Drug Discov. Devel. 5(2) (2002) 251-260.
[2] D.J. Craik, M. Cemazar, C.K. Wang, N.L. Daly, The cyclotide family of circular miniproteins: nature's combinatorial peptide template. Biopolymers 84(3) (2006) 250-266.
[3] K. Jagadish, J.A. Camarero, Cyclotides, a promising molecular scaffold for peptide-based therapeutics. Biopolymers 94(5) 611-616.
[4] A.E. Garcia, J.A. Camarero, Biological activities of natural and engineered cyclotides, a novel molecular scaffold for peptide-based therapeutics. Curr Mol Pharmacol 3(3) 153-163.
[5] N.L. Daly, K.J. Rosengren, D.J. Craik, Discovery, structure and biological activities of cyclotides. Adv Drug Deliv Rev 61(11) (2009) 918-930.
[6] O. Saether, D.J. Craik, I.D. Campbell, K. Sletten, J. Juul, D.G. Norman, Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide kalata B1. Biochemistry 34(13) (1995) 4147-4158.
[7] L. Cascales, D.J. Craik, Naturally occurring circular proteins: distribution, biosynthesis and evolution. Org Biomol Chem 8(22) (2010) 5035-5047.
[8] A.G. Poth, M.L. Colgrave, R. Philip, B. Kerenga, N.L. Daly, M.A. Anderson, D.J. Craik, Discovery of cyclotides in the fabaceae plant family provides new insights into the cyclization, evolution, and distribution of circular proteins. ACS Chem Biol 6(4) (2011) 345-355.
[9] J.P. Mulvenna, C. Wang, D.J. Craik, CyBase: a database of cyclic protein sequence and structure. Nucleic Acids Res 34(Database issue) (2006) D192-194.
[10] C.K. Wang, Q. Kaas, L. Chiche, D.J. Craik, CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering. Nucleic Acids Res 36(Database issue) (2008) D206-210.
[11] J.P. Mulvenna, J.S. Mylne, R. Bharathi, R.A. Burton, N.J. Shirley, G.B. Fincher, M.A. Anderson, D.J. Craik, Discovery of cyclotide-like protein sequences in graminaceous crop plants: ancestral precursors of circular proteins? Plant Cell 18(9) (2006) 2134-2144.
[12] C.W. Gruber, A.G. Elliott, D.C. Ireland, P.G. Delprete, S. Dessein, U. Goransson, M. Trabi, C.K. Wang, A.B. Kinghorn, E. Robbrecht, D.J. Craik, Distribution and evolution of circular miniproteins in flowering plants. Plant Cell 20(9) (2008) 2471-2483.
[13] I. Saska, A.D. Gillon, N. Hatsugai, R.G. Dietzgen, I. Hara-Nishimura, M.A. Anderson, D.J. Craik, An asparaginyl endopeptidase mediates in vivo protein backbone cyclization. The Journal of biological chemistry 282(40) (2007) 29721-29728.
[14] A.D. Gillon, I. Saska, C.V. Jennings, R.F. Guarino, D.J. Craik, M.A. Anderson, Biosynthesis of circular proteins in plants. Plant J 53(3) (2008) 505-515.
[15] B.F. Conlan, A.D. Gillon, D.J. Craik, M.A. Anderson, Circular proteins and mechanisms of cyclization. Biopolymers 94(5) (2010) 573-583.
[16] N.L. Daly, S. Love, P.F. Alewood, D.J. Craik, Chemical synthesis and folding pathways of large cyclic polypeptides: studies of the cystine knot polypeptide kalata B1. Biochemistry 38(32) (1999) 10606-10614.
[17] J.P. Tam, Y.A. Lu, J.L. Yang, K.W. Chiu, An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. Proc Natl Acad Sci U S A 96(16) (1999) 8913-8918.
[18] P. Thongyoo, E.W. Tate, R.J. Leatherbarrow, Total synthesis of the macrocyclic cysteine knot microprotein MCoTI-II. Chem Commun (Camb)(27) (2006) 2848-2850.
[19] P. Thongyoo, N. Roque-Rosell, R.J. Leatherbarrow, E.W. Tate, Chemical and biomimetic total syntheses of natural and engineered MCoTI cyclotides. Org Biomol Chem 6(8) (2008) 1462-1470.
[20] R.H. Kimura, A.T. Tran, J.A. Camarero, Biosynthesis of the cyclotide kalata B1 by using protein splicing. Angew Chem Int Ed 45(6) (2006) 973-976.
[21] J.A. Camarero, R.H. Kimura, Y.H. Woo, A. Shekhtman, J. Cantor, Biosynthesis of a fully functional cyclotide inside living bacterial cells. Chembiochem 8(12) (2007) 1363-1366.

[22] J. Austin, W. Wang, S. Puttamadappa, A. Shekhtman, J.A. Camarero, Biosynthesis and biological screening of a genetically encoded library based on the cyclotide MCoTI-I. Chembiochem 10(16) (2009) 2663-2670.
[23] J. Austin, R.H. Kimura, Y.H. Woo, J.A. Camarero, In vivo biosynthesis of an Ala-scan library based on the cyclic peptide SFTI-1. Amino Acids 38(5) (2010) 1313-1322.
[24] N.L. Daly, K.R. Gustafson, D.J. Craik, The role of the cyclic peptide backbone in the anti-HIV activity of the cyclotide kalata B1. FEBS Lett 574(1-3) (2004) 69-72.
[25] J.F. Hernandez, J. Gagnon, L. Chiche, T.M. Nguyen, J.P. Andrieu, A. Heitz, T. Trinh Hong, T.T. Pham, D. Le Nguyen, Squash trypsin inhibitors from Momordica cochinchinensis exhibit an atypical macrocyclic structure. Biochemistry 39(19) (2000) 5722-5730.
[26] A. Heitz, J.F. Hernandez, J. Gagnon, T.T. Hong, T.T. Pham, T.M. Nguyen, D. Le-Nguyen, L. Chiche, Solution structure of the squash trypsin inhibitor MCoTI-II. A new family for cyclic knottins. Biochemistry 40(27) (2001) 7973-7983.
[27] M.E. Felizmenio-Quimio, N.L. Daly, D.J. Craik, Circular proteins in plants: solution structure of a novel macrocyclic trypsin inhibitor from Momordica cochinchinensis. J Biol Chem 276(25) (2001) 22875-22882.
[28] R.J. Clark, N.L. Daly, D.J. Craik, Structural plasticity of the cyclic-cystine-knot framework: implications for biological activity and drug design. Biochem J 394(Pt 1) (2006) 85-93.
[29] D.J. Craik, M. Cemazar, N.L. Daly, The cyclotides and related macrocyclic peptides as scaffolds in drug design. Curr Opin Drug Discov Devel 9(2) (2006) 251-260.
[30] D.J. Craik, N.L. Daly, J. Mulvenna, M.R. Plan, M. Trabi, Discovery, structure and biological activities of the cyclotides. Curr Protein Pept Sci 5(5) (2004) 297-315.
[31] S. Reiss, M. Sieber, V. Oberle, A. Wentzel, P. Spangenberg, R. Claus, H. Kolmar, W. Losche, Inhibition of platelet aggregation by grafting RGD and KGD sequences on the structural scaffold of small disulfide-rich proteins. Platelets 17(3) (2006) 153-157.
[32] H. Kolmar, Biological diversity and therapeutic potential of natural and engineered cystine knot miniproteins. Curr Opin Pharmacol 9(5) (2009) 608-614.
[33] S.S. Puttamadappa, K. Jagadish, A. Shekhtman, J.A. Camarero, Backbone Dynamics of Cyclotide MCoTI-I Free and Complexed with Trypsin. Angew Chem Int Ed Engl (2010).
[34] H. Sancheti, J.A. Camarero, "Splicing up" drug discovery. Cell-based expression and screening of genetically-encoded libraries of backbone-cyclized polypeptides. Adv Drug Deliv Rev 61(11) (2009) 908-917.
[35] K.P. Greenwood, N.L. Daly, D.L. Brown, J.L. Stow, D.J. Craik, The cyclic cystine knot miniprotein MCoTI-II is internalized into cells by macropinocytosis. Int J Biochem Cell Biol 39(12) (2007) 2252-2264.
[36] L.J. Hewlett, A.R. Prescott, C. Watts, The coated pit and macropinocytic pathways serve distinct endosome populations. J Cell Biol 124(5) (1994) 689-703.
[37] M.C. Kerr, R.D. Teasdale, Defining macropinocytosis. Traffic 10(4) (2009) 364-371.
[38] M. Lundberg, M. Johansson, Positively charged DNA-binding proteins cause apparent cell membrane translocation. Biochem Biophys Res Commun 291(2) (2002) 367-371.
[39] J.P. Richard, K. Melikov, E. Vives, C. Ramos, B. Verbeure, M.J. Gait, L.V. Chernomordik, B. Lebleu, Cell-penetrating peptides. A reevaluation of the mechanism of cellular uptake. J Biol Chem 278(1) (2003) 585-590.
[40] J.P. Lim, P.A. Gleeson, Macropinocytosis: an endocytic pathway for internalising large gulps. Immunol Cell Biol (2011).
[41] C.C. Norbury, L.J. Hewlett, A.R. Prescott, N. Shastri, C. Watts, Class I MHC presentation of exogenous soluble antigen via macropinocytosis in bone marrow macrophages. Immunity 3(6) (1995) 783-791.
[42] R.M. Steinman, S.E. Brodie, Z.A. Cohn, Membrane flow during pinocytosis. A stereologic analysis. J Cell Biol 68(3) (1976) 665-687.
[43] S.D. Conner, S.L. Schmid, Regulated portals of entry into the cell. Nature 422(6927) (2003) 37-44.

[44] G.J. Doherty, H.T. McMahon, Mechanisms of endocytosis. Annu Rev Biochem 78 (2009) 857-902.
[45] C. Kaduk, H. Wenschuh, M. Beyermann, K. Forner, L.A. Carpino, M. Bienert, Synthesis of Fmoc-amino acid fluorides via DAST, an alternative fluorinating agent Lett. Pept. Sci. 2(5) (1997) 285-288.
[46] R. Ingenito, D. Dreznjak, S. Guffler, H. Wenschuh, Efficient loading of sulfonamide safety-catch linkers by Fmoc amino acid fluorides. Org. Lett. 4(7) (2002) 1187-1188.
[47] J.A. Camarero, A.R. Mitchell, Synthesis of proteins by native chemical ligation using Fmoc-based chemistry. Protein Pept Lett 12(8) (2005) 723-728.
[48] R. Ingenito, E. Bianchi, D. Fattori, A. Pessi, Solid phase synthesis of peptide C-terminal thioesters by Fmoc/t-Bu chemistry
Source. J. Am. Chem. Soc. 121(49) (1999) 11369-11374.
[49] Y. Shin, K.A. Winans, B.J. Backes, S.B.H. Kent, J.A. Ellman, C.R. Bertozzi, Fmoc-Based Synthesis of Peptide-^{a}Thioesters: Application to the Total Chemical Synthesis of a Glycoprotein by Native Chemical Ligation. J. Am. Chem. Soc. 121 (1999) 11684-11689.
[50] J.A. Camarero, T.W. Muir, Chemoselective backbone cyclization of unprotected peptides. J. Chem. Soc., Chem. Comm. 1997 (1997) 1369-1370.
[51] L. Zhang, J.P. Tam, Synthesis and application of unprotected cyclic peptides as building blocks for peptide dendrimers. J. Am. Chem. Soc. 119 (1997) 2363-2370.
[52] J.A. Camarero, J. Pavel, T.W. Muir, Chemical Synthesis of a Circular Protein Domain: Evidence for Folding-Assisted Cyclization. Angew. Chem. Int. Ed. 37(3) (1998) 347-349.
[53] Y. Shao, W.Y. Lu, S.B.H. Kent, A novel method to synthesize cyclic peptides. Tetrahedron Lett. 39(23) (1998) 3911-3914.
[54] J.A. Camarero, T.W. Muir, Biosynthesis of a Head-to-Tail Cyclized Protein with Improved Biological Activity. J. Am. Chem. Soc. 121 (1999) 5597-5598.
[55] J.A. Camarero, D. Fushman, D. Cowburn, T.W. Muir, Peptide chemical ligation inside living cells: in vivo generation of a circular protein domain. Bioorg Med Chem 9(9) (2001) 2479-2484.
[56] S.C. Silverstein, R.M. Steinman, Z.A. Cohn, Endocytosis. Annu Rev Biochem 46 (1977) 669-722.
[57] S.B. Fonseca, M.P. Pereira, S.O. Kelley, Recent advances in the use of cell-penetrating peptides for medical and biological applications. Adv Drug Deliv Rev 61(11) (2009) 953-964.
[58] F. Duchardt, M. Fotin-Mleczek, H. Schwarz, R. Fischer, R. Brock, A comprehensive model for the cellular uptake of cationic cell-penetrating peptides. Traffic 8(7) (2007) 848-866.
[59] D. Brown, I. Sabolic, Endosomal pathways for water channel and proton pump recycling in kidney epithelial cells. J Cell Sci Suppl 17 (1993) 49-59.
[60] W. Shurety, N.L. Stewart, J.L. Stow, Fluid-phase markers in the basolateral endocytic pathway accumulate in response to the actin assembly-promoting drug Jasplakinolide. Mol Biol Cell 9(4) (1998) 957-975.
[61] K. Thompson, M.J. Rogers, F.P. Coxon, J.C. Crockett, Cytosolic entry of bisphosphonate drugs requires acidification of vesicles after fluid-phase endocytosis. Mol Pharmacol 69(5) (2006) 1624-1632.
[62] I.R. Nabi, P.U. Le, Caveolae/raft-dependent endocytosis. J Cell Biol 161(4) (2003) 673-677.
[63] F. Huang, A. Khvorova, W. Marshall, A. Sorkin, Analysis of clathrin-mediated endocytosis of epidermal growth factor receptor by RNA interference. J Biol Chem 279(16) (2004) 16657-16661.
[64] X. Jiang, F. Huang, A. Marusyk, A. Sorkin, Grb2 regulates internalization of EGF receptors through clathrin-coated pits. Mol Biol Cell 14(3) (2003) 858-870.
[65] A. Sorkin, M. Von Zastrow, Signal transduction and endocytosis: close encounters of many kinds. Nat Rev Mol Cell Biol 3(8) (2002) 600-614.
[66] C.J. Burckhardt, U.F. Greber, Virus movements on the plasma membrane support infection and transmission between cells. PLoS Pathog 5(11) (2009) e1000621.

[67] G.V. Jerdeva, K. Wu, F.A. Yarber, C.J. Rhodes, D. Kalman, J.E. Schechter, S.F. Hamm-Alvarez, Actin and non-muscle myosin II facilitate apical exocytosis of tear proteins in rabbit lacrimal acinar epithelial cells. J Cell Sci 118(Pt 20) (2005) 4797-4812.
[68] J. Mercer, A. Helenius, Virus entry by macropinocytosis. Nat Cell Biol 11(5) (2009) 510-520.
[69] J.R. Peterson, T.J. Mitchison, Small molecules, big impact: a history of chemical inhibitors and the cytoskeleton. Chem Biol 9(12) (2002) 1275-1285.
[70] P.U. Le, I.R. Nabi, Distinct caveolae-mediated endocytic pathways target the Golgi apparatus and the endoplasmic reticulum. J Cell Sci 116(Pt 6) (2003) 1059-1071.
[71] R.G. Parton, K. Simons, The multiple faces of caveolae. Nat Rev Mol Cell Biol 8(3) (2007) 185-194.
[72] A. Spang, On the fate of early endosomes. Biol Chem 390(8) (2009) 753-759.
[73] E.L. Eskelinen, Y. Tanaka, P. Saftig, At the acidic edge: emerging functions for lysosomal membrane proteins. Trends Cell Biol 13(3) (2003) 137-145.
[74] M. Fukuda, Lysosomal membrane glycoproteins. Structure, biosynthesis, and intracellular trafficking. J Biol Chem 266(32) (1991) 21327-21330.
[75] M.N. Cordonnier, D. Dauzonne, D. Louvard, E. Coudrier, Actin filaments and myosin I alpha cooperate with microtubules for the movement of lysosomes. Mol Biol Cell 12(12) (2001) 4013-4029.
[76] R. Matteoni, T.E. Kreis, Translocation and clustering of endosomes and lysosomes depends on microtubules. J Cell Biol 105(3) (1987) 1253-1265.
[77] J. Taunton, B.A. Rowning, M.L. Coughlin, M. Wu, R.T. Moon, T.J. Mitchison, C.A. Larabell, Actin-dependent propulsion of endosomes and lysosomes by recruitment of N-WASP. J Cell Biol 148(3) (2000) 519-530.
[78] F. Aniento, N. Emans, G. Griffiths, J. Gruenberg, Cytoplasmic dynein-dependent vesicular transport from early to late endosomes. J Cell Biol 123(6 Pt 1) (1993) 1373-1387.
[79] S. Loubery, C. Wilhelm, I. Hurbain, S. Neveu, D. Louvard, E. Coudrier, Different microtubule motors move early and late endocytic compartments. Traffic 9(4) (2008) 492-509.
[80] A.D. Frankel, C.O. Pabo, Cellular uptake of the tat protein from human immunodeficiency virus. Cell 55(6) (1988) 1189-1193.
[81] J.S. Lee, Q. Li, J.Y. Lee, S.H. Lee, J.H. Jeong, H.R. Lee, H. Chang, F.C. Zhou, S.J. Gao, C. Liang, J.U. Jung, FLIP-mediated autophagy regulation in cell death control. Nat Cell Biol 11(11) (2009) 1355-1362.
[82] K.J. Rosengren, N.L. Daly, M.R. Plan, C. Waine, D.J. Craik, Twists, knots, and rings in proteins. Structural definition of the cyclotide framework. J Biol Chem 278(10) (2003) 8606-8616.

Cyclotides from the pea plant family reveal the genetic origin for knotted circular proteins

Julio A. Camarero[1]

Department of Pharmacology and Pharmaceutical Sciences, School of Pharmacy, University of Southern California, Los Angeles, CA 90033

Cyclotides are fascinating circular proteins ranging from 28 to 37 aa residues that are naturally expressed in plants. They exhibit antimicrobial, insecticidal, antihelmintic, cytotoxic, and antiviral activities (1), and protease inhibitory activity (2), and can exert uterotonic effects (3). They all share a unique head-to-tail circular knotted topology of three disulfide bridges, with one disulfide bond penetrating through a macrocycle formed by the other two disulfides bonds and interconnecting peptide backbones, forming what is called a cystine knot topology (Fig. 1) (1). This cyclic cystine knot framework gives cyclotides a compact, highly rigid structure (4), which confers exceptional resistance to thermal/chemical denaturation and enzymatic degradation (5), thereby making cyclotides a promising molecular scaffold for drug discovery (6, 7). So far, cyclotides have been discovered in plants from the Rubiaceae (coffee), Violaceae (violet), and Cucurbitaceae (squash) families (8, 9), and more recently in the Fabaceae (legume) family (Fig. 1) (10). The discovery of cyclotides in the Fabaceae family of plants represents an important new development because this family of plants is the third largest on Earth, comprising approximately 18,000 different species. Some of these species are widely used as crops in human nutrition and food supply. This opens the intriguing possibility of using these plants for the large-scale production of cyclotides with pharmaceutical or agrochemical properties by using transgenic crops. The key to accomplishing that, however, is to have a better understanding of the mechanism that produces these interesting microproteins in this family of plants.

The report by Poth et al. in PNAS (11) brings us closer to that exciting possibility by describing the gene encoding the protein precursor of a unique cyclotide (Cter M) isolated from the leaf of butterfly pea (*Clitoria ternatea*), a representative member of the Fabaceae plant family. All the cyclotides reported so far from the Violaceae and Rubiaceae families are biosynthesized via processing from dedicated genes that, in some cases, encode multiple copies of the same cyclotide, and in others, mixtures of different cyclotide sequences (Fig. 1) (12). Poth et al. (11) reveal that the sequence encoding the cyclotide Cter M, however, is embedded within the albumin-1 gene of *C. ternatea* (Fig. 1). Plant albumins are part of the nutrient reservoir, but they also play a role in host defense. Generic albumin-1 genes are comprised of an ER signal sequence followed by an albumin chain-b, a linker, and an albumin chain-a. In the precursor of cyclotide Cter M, the cyclotide domain replaces the albumin chain-b domain. This interesting finding raises the question of how this replacement took place in evolution. There are two possibilities: (*i*) gradual evolution of the chain-b domain into the cyclotide domain or (*ii*) rapid lateral transfer of the cyclotide gene into the albumin gene. Poth et al. (11) present evidence supporting a gradual evolutionary path, whereby the albumin-1 chain-b slowly evolved into a more stable cyclotide domain. For example, the pea albumin-1 subunit-b (PA1b), one of the best-studied Fabaceae albumin components, is a 37-aa peptide from pea seeds (*Pisum sativum*), which also contains a cystine-knot structure (13). Remarkably, the cystine-knot core of PA1b overlays extremely well with that of the cyclotide Cter M. The composition and size of the PA1b loops are, however, totally different from those of the cyclotide Cter M. Recent mutagenesis studies on PA1b have also recently shown that this albumin domain is highly tolerant to mutations outside the cystine knot core (14). These observations support the possibility of divergent evolution of cyclotides from ancestral albumin domains, wherein evolution and natural selection provided an alternative loop decoration of the original cystine knot albumin core.

A final question remains. Linear cystine knot proteins such as PA1b are not backbone

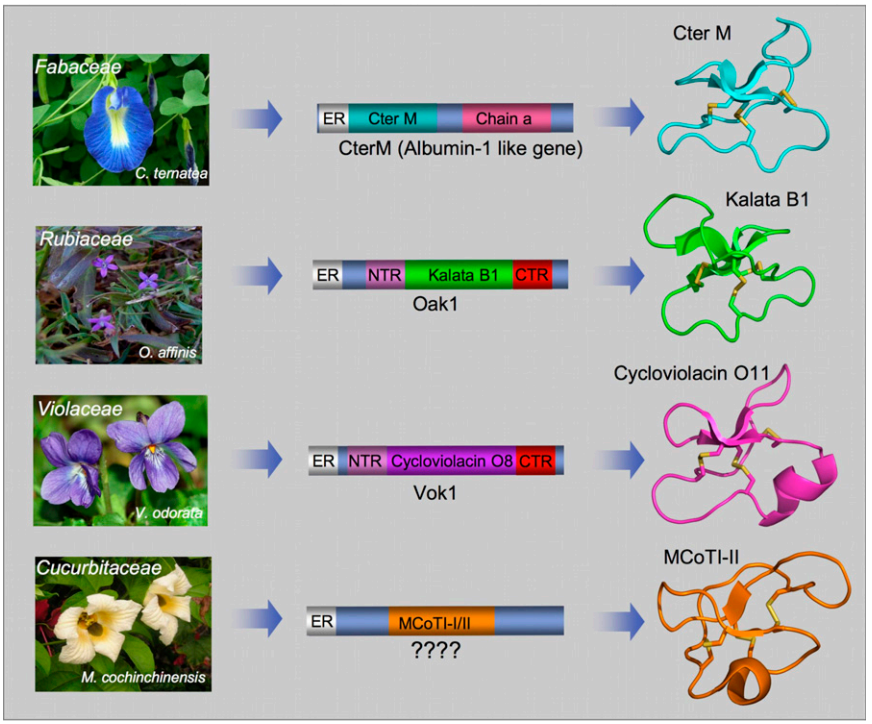


Fig. 1. Genetic origin of cyclotides from different plant families. Rubiaceae (*Oldenlandia affinis*) and Violaceae (*Viola odorata*) plants have dedicated genes for the production of cyclotides (12). These cyclotide precursors comprise an ER signal peptide, an N-terminal Pro region, the N-terminal repeat (NTR), the mature cyclotide domain, and a C-terminal flanking region (CTR). In contrast, the *CterM* gene (*C. ternatea*, Fabaceae) shows an ER signal peptide immediately followed by the cyclotide domain, which is flanked at the C terminus by a peptide linker and the albumin a-chain. The Cter M cyclotide domain replaces albumin-1 b-chain. The genetic origin of the *Cucurbitaceae* cyclotides (found in the seeds of *M. cochinchinensis*) remain to be identified.

Author contributions: J.A.C. wrote the paper.

The authors declare no conflict of interest.

See companion article 10.1073/pnas.1103660108.

[1]E-mail: jcamarer@usc.edu.

cyclized like cyclotides. What made the cyclization process possible, allowing the final transformation of an evolved cystine-knot albumin domain into a cyclotide? As indicated by Poth et al. (11), the structural analysis of PA1b may reveal some clues about how this could have happened. The NMR structure of PA1b (13) reveals that the N- and C-termini are very close to each other, and it is possible that mutations in the albumin genes predisposed them to cyclization during the evolution process.

However, what type of mutations could allow the backbone cyclization of a linear cystine knot albumin domain? Although the complete mechanism of how cyclotide precursors are processed and cyclized is not been fully characterized yet, recent studies indicate that an asparaginyl endopeptidase (AEP; also known as vacuolar processing enzyme or legumain) is a key element in the cyclization of cyclotides (15, 16). It has been proposed that the cyclization step mediated by AEP takes place at the same time as the cleavage of the C-terminal propeptide from the cyclotide precursor protein through a transpeptidation reaction (15). The transpeptidation reaction involves an acyl-transfer step from the acyl-AEP intermediate to the N-terminal residue of the cyclotide domain (16). A similar process has been used for the chemical (17), chemoenzymatic (18), and recombinant (19) production of cyclotides. AEPs are Cys proteases that are very common in plants and are able to specifically cleave the peptide bond at the C terminus of Asn and, less efficiently, Asp residues. All the cyclotide precursors identified so far, including those from *C. ternatea*, contain a well conserved Asn/Asp residue at the C terminus of the cyclotide domain, which is consistent with the idea that cyclotides are cyclized by a transpeptidation reaction mediated by AEP (15).

Despite these similarities, *C. ternatea* cyclotides also show some differences regarding the residue immediately following the mechanistically conserved Asn. In the cyclotide precursors from the Violaceae and Rubiaceae families, the C-terminal Asn/Asp residue is always followed by a small amino acid, either Gly or Ser. However, the Cter M precursor reported by Poth et al. (11) indicates that a small amino acid is not always required in that position. Moreover, some *C. ternatea* cyclotides also have a His residue at the N terminus of the cyclotide precursor rather than the most common Gly residue found in most cyclotide domains (10). These observations seem to indicate that, at least in the Fabaceae family, the AEP-mediated transpeptidation step may be more tolerant than previously recognized.

The finding that albumin genes can evolve into protein precursors that can be subsequently processed to become cyclic was described in a recent report on the biosynthesis of the sunflower trypsin inhibitor peptide, SFTI-1 (20). SFTI-1 is a 14-residue peptide isolated from sunflower seeds with a head-to-tail cyclic backbone structure having only a single disulfide bond. In this case, the SFTI-1 linear precursor is embedded within a "napin-type" 2S albumin.

The report by Poth et al. (11) indicates that the biosynthetic origin of some cyclotides are very different from others, which could suggest that cyclic peptides might be more widely distributed than is currently realized. The exceptional stability of backbone-cyclized peptides may give them an evolutionary advantage, which may provide the driving force for the evolution of multiple biosynthetic pathways including the use of dedicated or recycled genes, with albumins now being implicated in the biosynthesis of two different classes of cyclic peptides.

In this context, it is worth noting that the protein precursors of the only two cyclotides isolated so far from the Cucurbitaceae plant family, *Momordica cochinchinensis* trypsin inhibitor I and II (MCoTI-I/II; Fig. 1), remain yet to be identified. These cyclotides are found in the seeds of *M. cochinchinensis* (a tropical squash plant) and are potent trypsin inhibitors. MCoTI cyclotides do not share significant sequence homology with the other cyclotides beyond the presence of the three-cystine bridges that adopt a similar backbone-cyclic cystine-knot topology (Fig. 1) and are more related to linear cystine-knot squash trypsin inhibitors. In fact, an acyclic version of MCoTI-cyclotides (known as MCoTI-III) can also be found in the seeds of *M. cochinchinensis*. This situation, in which the cyclic and linear versions of the cys-knot protein coexist in the same organism, provides a unique opportunity to study the genetic origin and evolution of these interesting molecules. Identification of the protein precursors for the cyclic and linear versions of these cystine-knot trypsin inhibitors should provide a unique snapshot in the evolutionary process of plant cyclic cystine-knot proteins.

In summary, the work by Poth et al. (11) provides critical information on the origin, Q:8 evolution, and processing of cyclotides from a plant of the Fabaceae family. The discovery of unique cyclotides as well as other cy- Q:9 clic peptides from a wide range of plants is key to define and fully understand the different cyclization mechanisms used by plants. So far, the expression of cyclotides in transgenic plants has been attempted only in Arabidopsis and tobacco (15, 16), in which cyclotide expression is highly inefficient, giving rise to mostly acyclic or truncated proteins. The proven ability of *C. ternatea* to produce fully folded cyclotides seems to suggest that other species of the Fabaceae family could also be used for the production of cyclotides. Several members of this large family of plants are agricultural crops, which opens the intriguing possibility of generating genetically engineered crops for the large-scale production of cyclotides with useful pharmacological or agrochemical properties in the near future.

ACKNOWLEDGMENTS. This work was supported by National Institutes of Health Research Grant R01-GM090323 and Department of Defense Congressionally Directed Medical Research Program Grant PC09305.

1. Daly NL, Rosengren KJ, Craik DJ (2009) Discovery, structure and biological activities of cyclotides. *Adv Drug Deliv Rev* 61:918–930.
2. Avrutina O, et al. (2005) Trypsin inhibition by macrocyclic and open-chain variants of the squash inhibitor MCoTI-II. *Biol Chem* 386:1301–1306.
3. Gran L, Sandberg F, Sletten K (2000) Oldenlandia affinis (R&S) DC. A plant containing uteroactive peptides used in African traditional medicine. *J Ethnopharmacol* 70: 197–203.
4. Puttamadappa SS, Jagadish K, Shekhtman A, Camarero JA (2010) Backbone dynamics of cyclotide MCoTI-I free and complexed with trypsin. *Angew Chem Int Ed Engl* 49:7030–7034.
5. Colgrave ML, Craik DJ (2004) Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: The importance of the cyclic cystine knot. *Biochemistry* 43: 5965–5975.
6. Jagadish K, Camarero JA (2010) Cyclotides, a promising molecular scaffold for peptide-based therapeutics. *Biopolymers* 94:611–616.
7. Garcia AE, Camarero JA (2010) Biological activities of natural and engineered cyclotides, a novel molecular scaffold for peptide-based therapeutics. *Curr Mol Pharmacol* 3:153–163.
8. Gruber CW, et al. (2008) Distribution and evolution of circular miniproteins in flowering plants. *Plant Cell* 20: 2471–2483.
9. Chiche L, et al. (2004) Squash inhibitors: From structural motifs to macrocyclic knottins. *Curr Protein Pept Sci* 5:341–349.
10. Poth AG, et al. (2011) Discovery of cyclotides in the Fabaceae plant family provides new insights into the cyclization, evolution, and distribution of circular proteins. *ACS Chem Biol* 6:345–355.
11. Poth AG, et al. (2011) Discovery of an unusual biosynthetic origin for circular proteins in legumes. *Proc Natl Acad Sci USA*, 10.1073/pnas.1103660108.
12. Dutton JL, et al. (2004) Conserved structural and sequence elements implicated in the processing of gene-encoded circular proteins. *J Biol Chem* 279:46858–46867.
13. Jouvensal L, et al. (2003) PA1b, an insecticidal protein extracted from pea seeds (Pisum sativum): 1H-2-D NMR study and molecular modeling. *Biochemistry* 42:11915–11923.
14. Da Silva P, et al. (2010) Molecular requirements for the insecticidal activity of the plant peptide pea albumin 1 subunit b (PA1b). *J Biol Chem* 285:32689–32694.
15. Saska I, et al. (2007) An asparaginyl endopeptidase mediates in vivo protein backbone cyclization. *J Biol Chem* 282:29721–29728.
16. Gillon AD, et al. (2008) Biosynthesis of circular proteins in plants. *Plant J* 53:505–515.
17. Craik DJ, Conibear AC (2011) The chemistry of cyclotides. *J Org Chem*, in press.
18. Thongyoo P, Roqué-Rosell N, Leatherbarrow RJ, Tate EW (2008) Chemical and biomimetic total syntheses of natural and engineered MCoTI cyclotides. *Org Biomol Chem* 6:1462–1470.
19. Austin J, Wang W, Puttamadappa S, Shekhtman A, Camarero JA (2009) Biosynthesis and biological screening of a genetically encoded library based on the cyclotide MCoTI-I. *ChemBioChem* 10:2663–2670.
20. Mylne JS, et al. (2011) Albumins and their processing machinery are hijacked for cyclic peptides in sunflower. *Nat Chem Biol* 7:257–259.

Pharm Res
DOI 10.1007/s11095-010-0325-1

EXPERT REVIEW

Protein Microarrays: Novel Developments and Applications

Luis Berrade • Angie E. Garcia • Julio A. Camarero

Received: 22 September 2010 / Accepted: 8 November 2010

ABSTRACT Protein microarray technology possesses some of the greatest potential for providing direct information on protein function and potential drug targets. For example, functional protein microarrays are ideal tools suited for the mapping of biological pathways. They can be used to study most major types of interactions and enzymatic activities that take place in biochemical pathways and have been used for the analysis of simultaneous multiple biomolecular interactions involving protein-protein, protein-lipid, protein-DNA and protein-small molecule interactions. Because of this unique ability to analyze many kinds of molecular interactions en masse, the requirement of very small sample amount and the potential to be miniaturized and automated, protein microarrays are extremely well suited for protein profiling, drug discovery, drug target identification and clinical prognosis and diagnosis. The aim of this review is to summarize the most recent developments in the production, applications and analysis of protein microarrays.

KEY WORDS drug discovery · protein chips · protein immobilization · protein profiling · proteomics

INTRODUCTION

Protein microarray technology has made enormous progress in the last decade, increasingly becoming an important research tool for the study and detection of proteins, protein-protein interactions and numerous other biotechnological applications (1–4). The use of protein microarrays has advantages over more traditional methods for the study of molecular interactions. They require low sample consumption and have potential for miniaturization. Protein microarrays displaying multiple biologically active proteins simultaneously have the potential to provide high-throughput protein analysis in the same way DNA arrays did for genomics research a decade ago. This is a feature that is extremely important for the analysis of protein interactions at the proteome-scale. The transition from DNA to protein microarrays, however, has required the development of specially tailored protein immobilization methods that ensure the protein structure and biological function after the immobilization step. Several technologies have been developed in the last few years that allow the site-specific immobilization of proteins onto solid supports for the rapid production of protein microarrays using high throughput expression systems, such as cell-free expression systems (5–7). The development of appropriate detection systems to monitor protein interactions has also been an important challenge for the optimal use of protein microarrays. The use of techniques such as fluorescence imaging, mass-spectrometry (MS) and surface plasmon resonance (SPR) were recently developed and adapted to be interfaced with protein micro-arrays. During the last decade, a number of excellent reviews have appeared in the literature describing the concept, preparation, analysis and applications of protein microarrays, highlighting the increasing importance of this technology (1–4,8). The aim of this review is to summarize the latest developments in protein microarray technology in the areas of protein immobilization, novel protein detection schemes and applications of this promising technology.

PROTEIN MICROARRAYS

Protein microarrays are usually divided in two groups: functional protein microarrays and protein-detecting

L. Berrade · A. E. Garcia · J. A. Camarero (✉)
Department of Pharmacology and Pharmaceutical Sciences
School of Pharmacy, University of Southern California
1985 Zonal Avenue, PSC 616
Los Angeles, California 90033, USA
e-mail: jcamarer@usc.edu

Published online: 30 November 2010

microarrays (Fig. 1) (2,9). Protein function microarrays are made by the immobilization of different purified proteins, protein domains or functional peptides. These types of microarrays are generally used to study molecular interactions and screen potential interacting partners. On the other hand, protein-detecting microarrays are made by the immobilization of specific protein capture reagents that can specifically recognize particular proteins from complex mixtures. These microarrays are used for protein profiling, i.e. quantification of protein abundances and evaluation of post-translational modifications in complex mixtures.

Functional Protein Microarrays

Understanding the network of molecular interactions that defines a particular proteome is one of the main goals of functional proteomics. Functional protein microarrays provide an extremely powerful tool to accomplish this daunting task, especially when assessing the activity of families of related proteins. In 2000, Schreiber and co-workers showed that purified recombinant proteins could be microarrayed onto chemically derivatized glass slides without seriously affecting their molecular and functional integrity (10). More recently, Snyder and co-workers have been able to immobilize ≈5,800 proteins from *Sacharomyces cerivisiae* onto microscope glass slides (11). This protein chip was then probed with different phospholipids to identify several lipid-binding proteins. The same authors also used this proteome chip for the identification of substrates for 87 different protein kinases (12). Using this microarray data set in combination with protein-protein interaction and transcription factor binding data, the authors were able to reveal several novel regulatory modules in yeast (12). Using a similar approach, Dinesh-Kumar and co-workers were able to construct a protein microarray containing 2,158 unique *Arabidopsis thaliana* proteins. This array was used for the identification of 570 phosphorylation substrates of mitogen-activated protein kinases, which included several transcription factors involved in the regulation of development, host immune defense, and stress responses (13). The analysis of proteome-wide microarrays from yeast was also recently used to find unexpected non-chromatin substrates for the essential nucleosomal acetyl transferase of H4 (NuA4) complex (14). In this interesting work, the authors discovered that NuA4 is a natural substrate for the metabolic enzyme phosphoenolpyruvate carboxykinase and that its acetylation is critical for regulating the chronological lifespan of yeast (14). In another example, human proteome arrays were used for the detection of autoimmune response markers in several human cancers (15,16). Kirschner and co-workers have also used human proteome arrays to identify novel substrates of the anaphase-promoting complex (17). This was accomplished by probing the arrays with cell extracts that replicate the mitotic checkpoint and anaphase release and then probing the captured proteins with antibodies specific for detecting poly-ubiquitination (17). Functional protein microarrays have also been used to study families of interacting protein

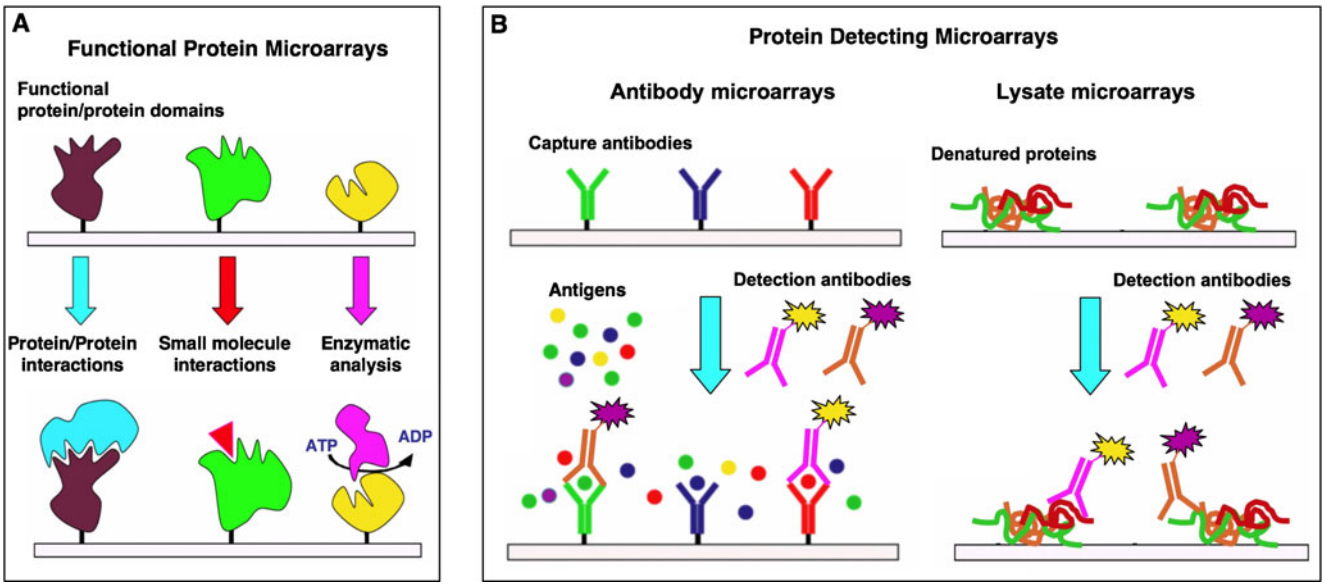


Fig. 1 Common formats used for the preparation of protein microarrays. Functional protein microarrays (**A**) are used to study and identify new molecular interactions between proteins, small molecules or enzyme substrates, for example. Protein detecting microarrays (**B**) are used to identify proteins from complex mixtures. In the sandwich format (**B**, left), captured proteins are detected by a secondary antibody typically labeled with a fluorescent dye to facilitate detection and quantification. In contrast to antibody microarrays, lysate microarrays (**B**, right) are typically immobilized onto nitrocellulose-coated glass slides (FAST slides) and detected using fluorescent-labeled solution-phase specific antibodies.

domains. Bedford and co-workers have shown that several protein domains (FF, FHA, PH, PDZ, SH2, SH3, and WW) can be immobilized onto a microarray format, retaining their ability to mediate specific interactions (18). Similar approaches were used to study the interactions associated with WW domains in yeast (19) and Kaposi-sarcoma viral proteins and the host endocytic machinery (20), and to evaluate the interactions between different proline-rich peptides derived from the myelin basic protein and several SH3 domains (21).

Functional protein domain microarrays can also be used to quantify protein interactions. For example, in 2004 Blackburn and co-workers used microarrays containing multiple variants of the transcription factor p53 to study and quantify their DNA-binding preferences (22). By using fluorescent-labeled DNA probes, the authors were able to produce binding isotherms and extract the different equilibrium dissociation constants for every p53 variant (22). MacBeath and co-workers have also used a similar approach to quantify the interactions of several human SH2 and PTB domains with different phosphotyrosine-containing peptides derived from human ErbB receptors (Fig. 2) (23). This type of protein microarray provides a unique way to study the binding properties of complete families of proteins and/or protein domains in an unbiased way. In addition, they have the potential to generate data that, when collected in a quantitative way, could be used for training predictive models of molecular recognition (24–26). As a recent example, MacBeath and co-workers recently used functional microarrays containing multiple murine PDZ protein domains to screen potential interactions with 217 genome-encoded peptides derived from the murine proteome (24,25). The data generated was used to train a multidomain selectivity model that was able to predict PDZ domain-peptide interactions across the mouse proteome. Interestingly, the models showed that PDZ domains are not grouped into discrete functional classes; instead, they are uniformly distributed throughout the selectivity space. This finding strongly suggests that the PDZ domains across the proteome are optimized to minimize cross-reactivity (24,25).

Protein-Detecting Microarrays

As described above, functional protein microarrays allow high-throughput screening and quantification of protein interactions on a proteome-wide scale, thus providing an unbiased perspective on the connectivity of the different protein-protein interaction networks. Establishing how this information flows through these interacting networks, however, requires measuring the abundance and post-translational modifications of many proteins from complex biological mixtures. Protein-detecting microarrays are ideal reagents for this type of analysis. One of the most frequently used strategies to prepare this type of microarray involves the use of monoclonal antibodies as specific protein capture reagents. Antibodies have been classically well suited for this task, since there are a large number of commercially available specific antibodies, which can be easily immobilized onto solid supports (4,27–30). However, the potential problems associated with the use of antibodies for chip assembly, which might manifest themselves through moderate expression yields and by issues related to the stability and solubility of these large proteins, have led to the exploration of alternative protein scaffolds as a source for new, more effective and stable protein capture reagents (24,31,32). Suitable protein scaffolds that have been proposed include fibronectin domains, the Z domain of protein A, lipocalins and cyclotides, among others.

In general, antibody microarrays are well suited for detecting changes in the abundances of proteins in biological samples with a relatively large dynamic range (33). For example, Haab and co-workers made use of antibody microarrays for serum-protein profiling in order to identify potential biomarkers in prostate cancer (33). Using this approach, the authors were able to identify five proteins (immunoglobulins G and M, α1-anti-chymotrypsin, villin and the Von Willebrand factor) that had significantly different levels of expression between the prostate cancer samples and control samples from healthy individuals.

In a similar fashion to that of a sandwich ELISA assay, quite often, antibody microarrays make use of a second antibody directed towards a different epitope of the protein to be analyzed. This facilitates the detection and quantification of the corresponding analyte. This approach has been used for monitoring changes in the phosphorylation state of host proteins (34), including receptor tyrosine kinases (35), and for serum protein profiling to identify new biomarkers in prostate cancer (36) among other applications. The use of this approach is usually limited, however, by the availability of suitable antibodies that can be used for capture and detection. Moreover, the detection step requires the simultaneous use of multiple fluorescent-labeled antibodies, which may increase background signal as well as the risk of cross-reactive binding as the number of antibodies increases. A way to overcome this problem is to label the proteins in the biological sample to be analyzed using one or more fluorescent dyes (37). This approach allows one to perform ratiometric comparisons between different samples by using spectrally distinct fluorophores. This strategy has been employed for the discovery of molecular biomarkers in different types of human cancer (38–40). It should be highlighted, however, that non-specific chemical labeling of proteins introduces chemical modifications on their surface and, therefore, may alter antibody recognition and lead to false signals. Also, this

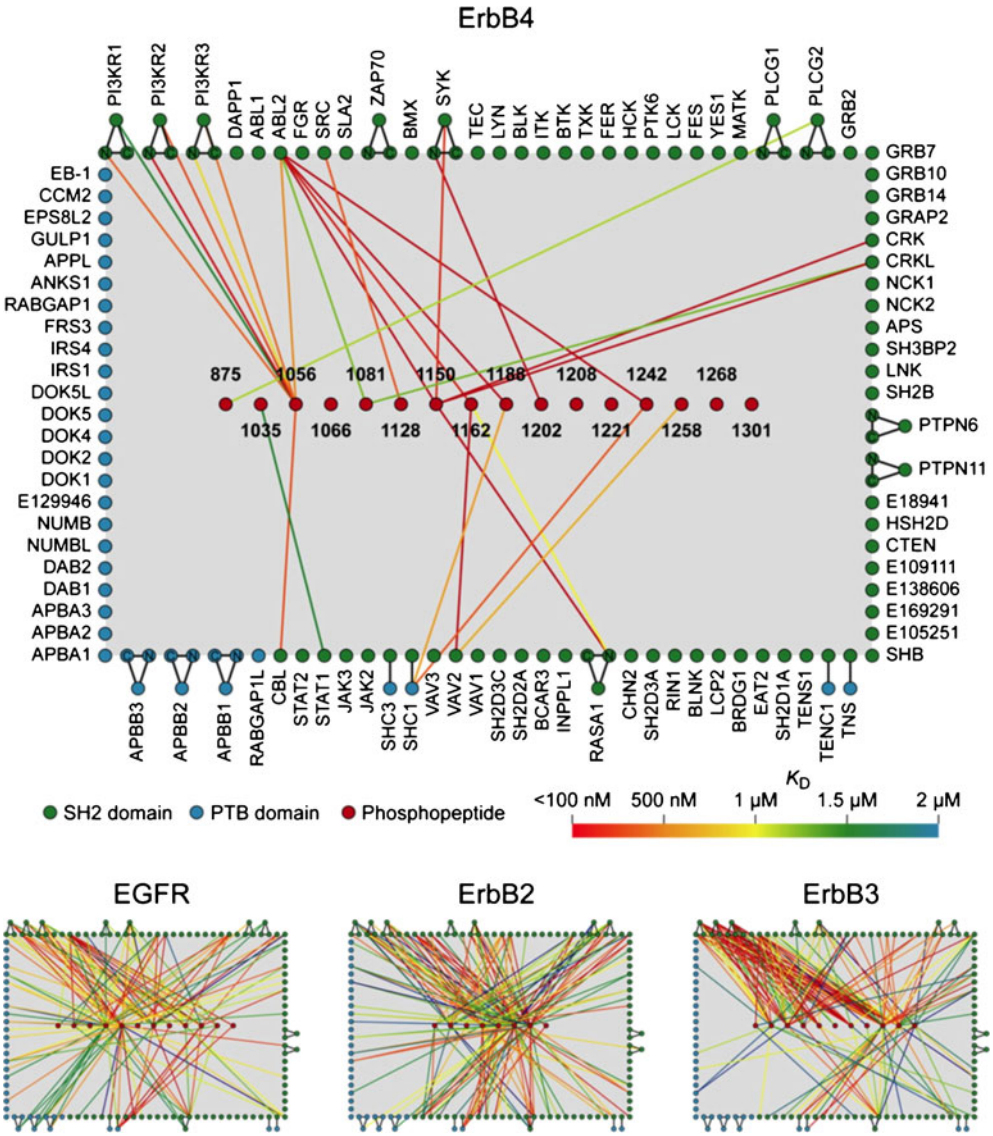


Fig. 2 Quantitative interaction networks of tyrosine kinases associated with the Erb family of receptors, which was determined using protein microarrays displaying 96 SH2 and 37 PTB domains. The SH2 and PTB protein domains were probed with fluorescently labeled phosphopeptides representing the different tyrosine phosphoryaltion sites on the Erb kinases. The readout of peptide binding was monitored and quantified by fluorescence. The interaction maps (bottom panel) were constructed from the quantitative interaction data (156). Reprinted from reference (156) with permission from Elsevier.

approach requires the homogeneous labeling of proteins across different samples, which in most cases cannot be completely guaranteed. These drawbacks can, in principle, be avoided by using a label-free detection scheme. However, nearly all of the different methods available for this task (see below) still lack the sensitivity required for most biological applications.

Although antibody microarrays are well suited for protein profiling, proteome-wide applications have not been accomplished yet. This is mainly due to the lack of available, well-validated antibodies. An ingenious solution proposed by Lauffenburger and co-workers, however, is to use a combination of different experimental approaches with the data generated by microarrays (41,42). In this work, the authors combined data gathered from antibody microarrays, enzymatic assays, immunoblotting, and flow cytometry to assemble a network of ≈10,000 interactions in HT-29 cells treated with different combinations of cytokines (41). All of this information was later used to uncover mechanisms of crosstalk involving pro- and anti-apoptotic signals induced by different cytokines (42).

Protein Lysate Microarrays

An interesting alternative to antibody microarrays is to immobilize cell lysates and then use specific monoclonal antibodies to identify and quantify the presence of a particular analyte in the corresponding lysate. This technology was first described by Liotta and co-workers to monitor pro-survival checkpoint proteins as a function of cancer progression (43). The same approach has recently been used for the discovery and validation of specific biomarkers for disease diagnosis and patient stratification. Utz and co-workers (44) have also made use of lysate microarrays to study the kinetics of intracellular signaling by tracking 62 phosphorylation sites in stimulated Jurkat

cells, which allowed them to discover a previously unknown connection between T-cell receptor activation and Raf-1 activity (44).

In protein lysate microarrays, every spot in the microarray contains the entire set of biological proteins to be analyzed. This means that in order to analyze the abundance and modification states of different proteins present in the lysate, it is necessary to prepare as many copies of the array as proteins needed to be analyzed. Lysate microarrays also denature the proteins to be analyzed during the immobilization step onto the solid support. This makes it impossible to study complex protein-protein interactions and requires the use of specific and well-validated antibodies for the recognition of specific continuous protein epitopes. This is a serious limitation of this technique, since it only allows the analysis of proteins that have already been discovered and to which suitable antibodies are available. In this regard, it should be noted the majority of commercially available antibodies typically show substantial cross-reactivity issues and, therefore, are not appropriate for this type of approach. Only antibodies able to provide a single band in a standard Western blot should be used. Moreover, the blocking and detection protocols, as well as the composition of the lysis buffer, have been shown to substantially affect antibody performance (45), therefore indicating that further developments are required for the widespread use of this technology.

NOVEL APPROACHES FOR PROTEIN IMMOBILIZATION

The immobilization of proteins onto solid supports has traditionally relied on non-specific adsorption (46,47) or covalent crosslinking of naturally occurring chemical groups within proteins (47–49). These approaches usually provide a random orientation of the immobilized protein onto the solid support, which may compromise the structural and/or functional integrity of the protein (50). This is a key issue for the fabrication of functional protein microarrays as described above. The use of recombinant affinity tags as capture reagents offers site-specific immobilization. The most commonly used affinity tags include biotin/avidin (51–53), His-tag/Ni^{2+}-nitriloacetic acid (11,54) and glutathione-S-transferase (GST)/glutathione (GSH) (12,55).

Immobilization of antibodies through the Fc region onto protein A- or protein G-coated surfaces has also been used for the creation of antibody microarrays (56,57). Additionally, thioredoxin (58), maltose-binding protein (59) and chitin-binding protein (60) have also been developed for the immobilization of the corresponding fusion proteins. Protein-DNA conjugates have also been recently reported for DNA-directed immobilization (DDI) of proteins onto complementary DNA-microarrays (61,62). Most of these interactions, however, are reversible and not stable over time (63–67). The use of site-specific chemical ligation reactions for the immobilization of proteins overcomes this limitation by allowing the proteins to be arranged in a defined, controlled fashion with exquisite chemical control (see references (29,68,69) for recent reviews in this field). This type of reaction requires two unique and mutually reactive groups on the protein and the solid support used for the immobilization step (Fig. 3). Ideally, the reaction between these groups should be highly chemoselective and compatible with physiological conditions to avoid denaturation during the immobilization step (28,70). Finally, it should be desirable that these unique reactive groups could be directly engineered into the proteins to be immobilized by using standard recombinant expression techniques.

Most of the chemoselective methods suitable for site-specific immobilization of proteins described in the literature rely on ligation methods originally designed for the chemical engineering of proteins (71–77). Key to these methods is the introduction of a unique reacting group at a defined position in the protein to be immobilized, which can later react in a chemoselective manner with a complementary group previously introduced into the surface (Fig. 3, see also references (4,27,29,69,78) for recent reviews).

Surface Modification

The most common solid supports employed for the immobilization of proteins in micro- and nano-biotechnology and biomedical applications involve the use of metals and silicon- and semiconductor-based substrates. Trialkoxysilanes such as 3-aminopropyl-trialkoxysilane (APS) or 3-mercaptopropyl-trialkoxysilane are typically employed for the chemical modification of silicon-based substrates for the introduction of amino ($–NH_2$) and thiol (–SH) groups, respectively. These chemical groups can then be modified by the introduction of appropriate linkers allowing the chemoselective attachment of proteins. Long chain alkyl-trichlorosilanes are more reactive towards the silanol group than trialkoxysilanes and have also been employed for the chemical modification of silicon-based substrates. The higher reactivity of long alkyl-trichlorosilanes is due to the self-assembling properties of the long aliphatic chains, which result in the formation of highly ordered and densely packed monolayers with solid-state-like properties (79,80).

Compounds containing the thiol or selenol (–SeH) groups can be also used to modify substrates based on transition metals, mostly gold and silver (80,81), or semiconductor materials (48). The chemical derivatization of gold surfaces using alkanethiols is by far one of the most commonly employed (81,82). Our group has developed

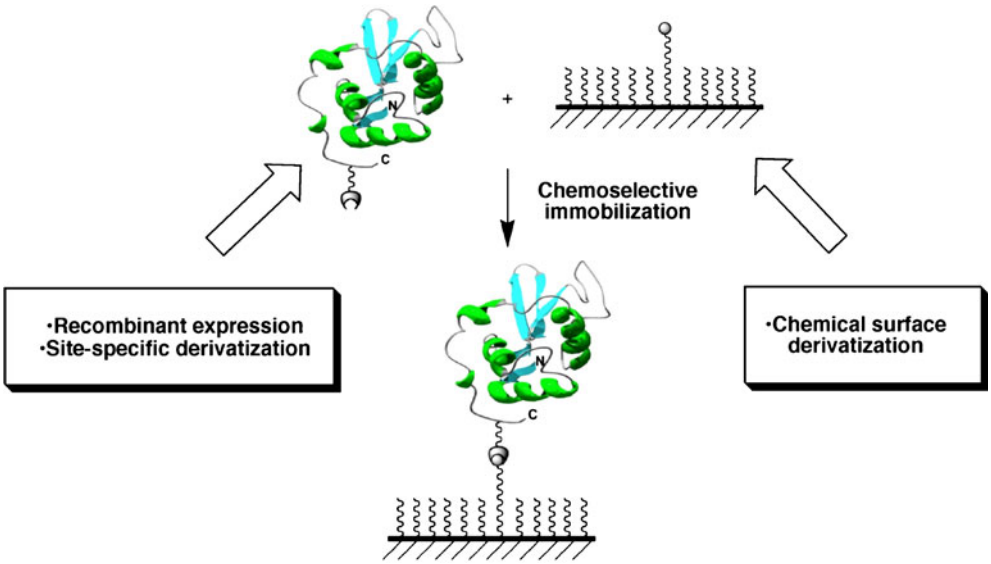


Fig. 3 Site-specific and covalent immobilization of a functional protein onto a chemically modified surface using a chemoselective ligation reaction.

several synthetic schemes for the efficient preparation of modified alkanethiols (83,84) that were used for the selective immobilization of functional proteins onto gold and glass surfaces (83–86).

The use of organic polymeric materials, such as polydimethylsiloxane (PDMS), poly-methylmethacrylate (PMMA) and polycarbonate (PC), has also been explored as a potential alternative to inorganic solid supports for the production of protein microarrays (87,88). The use of these materials also requires the introduction of suitable reacting groups for the site-specific immobilization of proteins. Common techniques usually employed for this task involve the use of plasma oxidation followed by treatment with appropriate organosilanes for the functionalization of PDMS (89), treatment of PMMA with 1,6-hexanediamine for the introduction of reactive amino groups (90), or using sulfonation reactions on PC to provide sulfated-coated surfaces (29).

Protein Immobilization Using Expressed Protein Ligation

The use of Expressed Protein Ligation (EPL) for the site-specific immobilization of biologically active proteins onto solid supports has been pioneered by our group (84). This approach relies on the chemoselective reaction of recombinantly produced protein α-thioesters with surfaces containing N-terminal Cys residues. C-terminal α-thioester proteins can be readily expressed in *Escherichia coli*, using commercially available intein expression systems (91). This ligation reaction is exquisitely chemoselective under physiological-like conditions and results in the site-specific immobilization of the protein through its C-terminus. We have successfully used this approach for the production of protein arrays containing several biologically active proteins onto Cys-coated glass slides (84). Typically, the immobilization reaction is performed at room temperature for 18 h and requires a minimal protein concentration in the low μM range for acceptable levels of immobilization (84). Yao and co-workers have also reported a similar approach for the selective immobilization of N-terminal Cys-containing polypeptides (52) and proteins (92) onto solid supports derivatized with an α-thioester group.

Schneider-Mergener and co-workers have recently combined SPOT synthesis (93) and a thioester ligation for the creation of arrays containing more than 10,000 variants of WW protein domains (94). Using 22 different peptide ligands to probe the WW domain arrays, the authors were able to monitor more than 250,000 binding experiments (94).

Protein Immobilization Using the Staudinger Ligation Reaction

A modified version of the Staudinger ligation reaction has also been employed for the chemoselective immobilization of azido-containing proteins onto solid supports derivatized with a suitable phosphine (71,75,95–97). The azido function can be readily incorporated into recombinant proteins using *E. coli* methionine auxotroph strains (98,99). A reactive arylphosphine derivative can be easily introduced onto carboxylic- or amine-containing surfaces (63,97). It should be noted that when the protein to be immobilized has multiple methionine residues this type of immobilization is not site-specific. This limitation can be overcome, however, by using *in vitro* EPL for the site-specific introduction of an azido group at the C-terminus of recombinant proteins (97). This can also be accomplished by reacting the corresponding protein C-terminal α-thioesters with functional hydrazines containing the azido

group for the site-specific introduction of this chemical group at the C-terminus of recombinant proteins (100).

Protein Immobilization Using "Click" Chemistry

The site-specific immobilization of azido- or alkyne-containing proteins onto alkyne- or azido-coated surfaces was recently accomplished by using the Cu(I)-catalyzed Huisgen 1,3-dipolar azide-alkyne cycloaddition, also known as "click" chemistry (101–103).

This is a very mild reaction that usually requires only the presence of Cu(I) as catalyst and is typically performed under physiological conditions. Under these conditions, the cycloaddition reaction is exquisitely regiospecific, affording only the 1,4-disubstitued tetrazole. The catalyst Cu(I) is usually generated *in situ* by reduction of Cu(II) using reducing agents such as tris-[2-carboxyethyl]-phosphine hydrochloride (TCEP•HCl) or ascorbic acid (77).

Site-specific incorporation of an alkyne group at the C-terminus of recombinant proteins can be also accomplished by using *in vitro* EPL (101) or nucleophilic cleavage of intein fusion proteins with derivatized hydrazines (100). The alkyne function has also been introduced chemo-enzymatically into recombinant proteins by using protein farnesyltransferases (PFTase) (102,103). This approach allows the selective S-alkylation of the Cys residue located in C-terminal Cys-Aaa-Aaa-Xxx motifs (where Xxx = Ala, Ser) by farnesyl diphosphate analogs containing the alkyne function.

Taki and co-workers have also accomplished the introduction of the azido function onto the N-termini of proteins by using the enzyme L/F-transferase (104), which is known to catalyze the transfer of hydrophobic amino acids from an aminoacyl-tRNA to the N-terminus (105). This modification, called NEXT-A (N-terminal extension of protein by transferase and amino-acyl transferase), can be accomplished in one pot and can also work in the presence of other proteins or even in crude protein mixtures (106,107). The authors used this method to functionalize the N-terminus of lectin EW29Ch with *p*-azido-phenylalanine, which was then immobilized onto a solid support coated with 4-dibenzocyclooctynol (DIB) through a copper-free "click" chemistry ligation (108–110).

Waldmann and co-workers have also developed the "click sulfonamide reaction" (CSR) between sulfonyl azides and alkynes to immobilize proteins and other types of biomolecules onto solid supports (111). Using this approach the authors were able to immobilize a C-terminal alkyne-modified Ras-binding domain (RBD) of cRaf1 onto a sulfonyl azide modified surface. The resulting immobilized protein was biologically active and able to selectively bind to GppNHp-bound Ras but not to inactive GDP-bound Ras (111).

In principle, "click" chemistry can be used for the chemoselective immobilization of alkyne- or azido-containing recombinant proteins onto azido- or alkyne-coated surfaces, respectively. However, it has been recently reported that the immobilization of alkyne-modified proteins onto azide-coated surfaces proceeds more efficiently (101). This effect could be attributed to the fact that the alkyne function coordinates Cu(I) in solution more efficiently than the azido group, which could improve the immobilization reaction (101). As for the other ligation reactions mentioned above, the minimal concentration of protein required for acceptable levels of immobilization using this type of ligation is typically found in the low μM range (101,102).

Protein Immobilization Using Active Site-Directed Capture Ligands

The efficiency of the different ligation reactions described so far for the site-specific immobilization of proteins onto solid supports depends strongly on the protein concentration in order to reach acceptable levels of immobilization (84,101,102). This intrinsic limitation could be in principle minimized by introducing two complementary interacting moieties on the protein and the surface, thus allowing the formation of a transient and specific intermolecular complex. The formation of this complex should be able to bring both reactive groups in close proximity, which would facilitate the efficiency of the ligation reaction (see Fig. 4). In this case, the efficiency of the reaction should not be dictated only by the concentration of the protein to be immobilized but rather by the affinity constant between the two interacting complementary moieties.

Mrksich and co-workers have used this approach for the selective immobilization of cutinase fusion proteins onto surfaces coated with chlorophosphonate ligands (112) (Fig. 5). Cutinase is a 22 kDa serine esterase, which can selectively react with chlorophosphonate ligands (113). These ligands bind with high affinity to the active site of the enzyme by mimicking the tetrahedral transition state stabilized by the esterase during the hydrolysis of the ester function. Once the complex is formed, the side-chain of the catalytic serine residue in the esterase active site reacts covalently with the chlorophosphonate group to form a relatively stable phosphate bond (Fig. 5). This approach was used for the immobilization of calmodulin (112) and for the preparation of antibody arrays (114) onto gold-coated self-assembled monolayers derivatized with a chlorophosphonate capture ligand.

Johnsson and co-workers have also used a similar approach for the site-specific immobilization of proteins but using human O^6-alkylguanine-DNA alkyltransferase (AGT) as a protein capture reagent (115). These types of enzymes can accept a benzyl group from O^6-benzylguanine (BG) deriva-

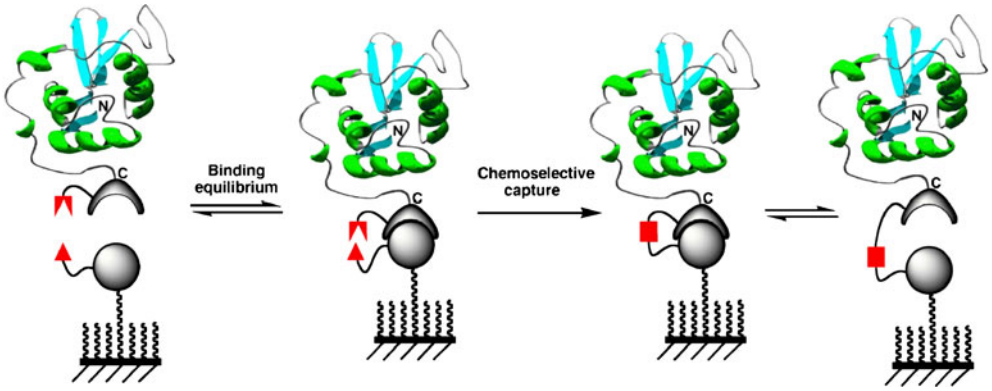


Fig. 4 Principle for site-specific protein immobilization using an active site-directed capture ligand approach.

tives, thus allowing the site-specific immobilization of AGT-fusion proteins onto O^6-benzylguanine-coated slides (116).

Protein Immobilization by Protein Trans-splicing

The main limitation of the site-specific capture methods described above is that they rely on the use of enzymes as capture reagents, which remain attached to the surface once the immobilization step is complete. The production of protein arrays containing these large linkers could give rise to non-specific interactions, especially in applications involving the analysis of complex samples (11,87).

Our group has recently developed a new traceless capture ligand approach for the site-specific attachment of proteins to surfaces based on the protein trans-splicing process (85) (Fig. 6). In protein trans-splicing, the intein self-processing domain is split in two fragments, which are referred as N-intein and C-intein (117,118). In this approach the N-intein fragment is fused to the C-terminus of the protein to be immobilized, and the C-intein fragment is immobilized onto

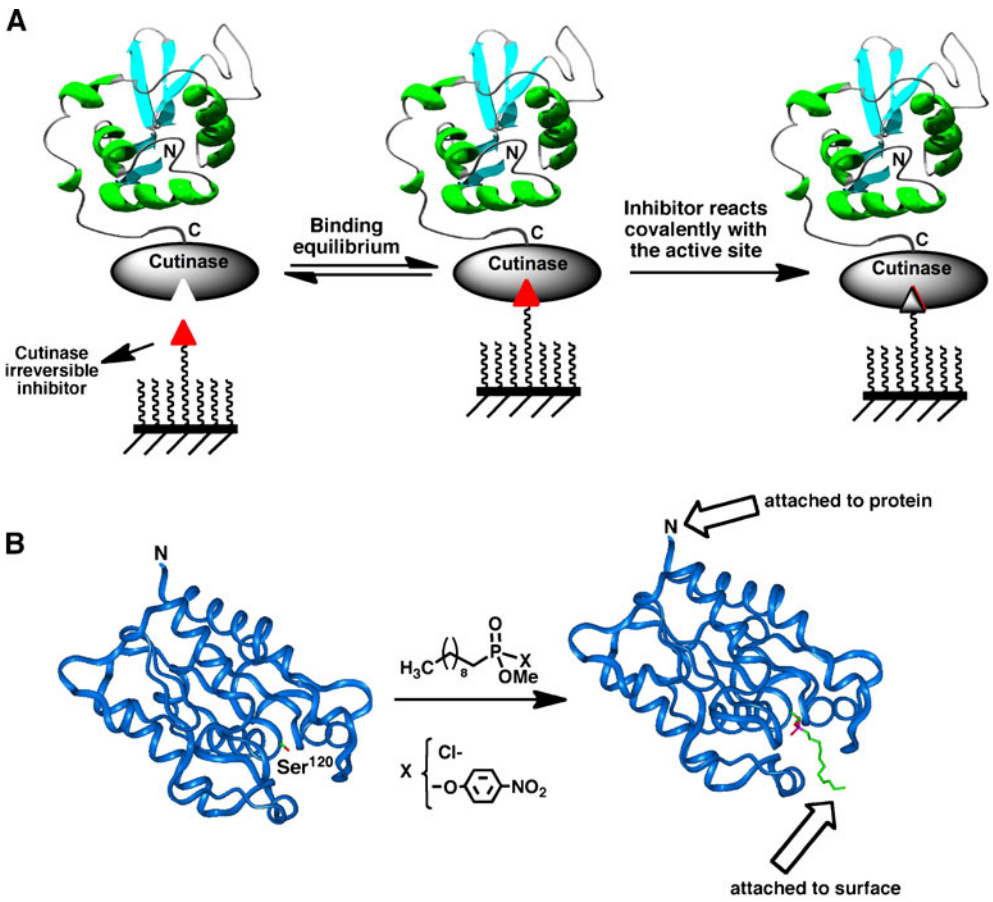


Fig. 5 **A** Site-specific immobilization of cutinase-fusion proteins using an active site-directed capture ligand. **B** Structure of *F. solani* cutinase enzyme free and bound to the inhibitor *n*-undecyl-O-methyl phosphonate chloride. The inhibitor is covalently bound through the side-chain hydroxyl group of the Ser^{120} residue, which is located at the active site of the enzyme (113).

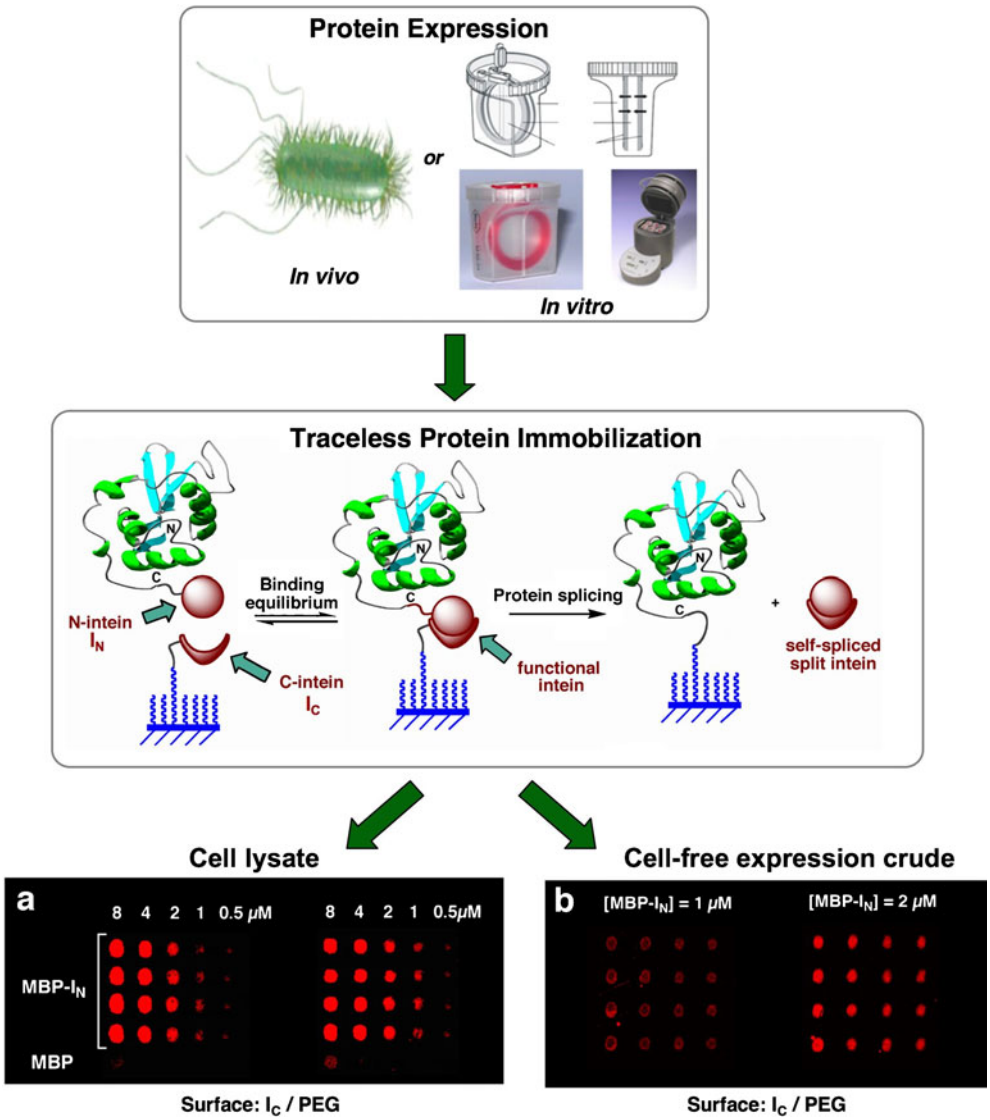


Fig. 6 Site-specific immobilization of proteins onto solid supports through protein trans-splicing (85). Maltose binding protein (MPB) was directly immobilized from (**a**) soluble cellular fraction of *E. coli* cells over-expressing MBP-I_N, and (**b**) MBP-I_N expressed *in vitro* using an *in vitro* trascription/traslation expression system. MBP was detected using a fluorescent-labeled specific antibody.

the solid support. When both intein fragments interact, they bind to each other with high affinity ($K_d \approx 200$ nM for the Ssp DnaE split-intein (85)), forming an active intein domain that can give rise to protein splicing in trans. This results in the immobilization of the protein of interest to the solid support at the same time that the split intein fragments are spliced out into solution (see Fig. 6). We have recently used this approach for the production of arrays containing several biologically active proteins onto chemically modified glass slides (85). The immobilization of proteins using trans-splicing is highly specific and efficient. For example, protein immobilization can be readily accomplished at concentrations in the low nM range (85). Importantly, the high specificity of protein trans-splicing allows the direct immobilization of proteins from complex mixtures, thus eliminating the need for purification and/or reconcentration of the proteins prior to the immobilization step. Furthermore, protein trans-splicing provides a completely traceless method of protein immobilization, since both intein fragments are spliced out into solution once the immobilization step is completed. Finally, protein trans-splicing was shown to be fully compatible with cell-free protein expression systems, which should facilitate high throughput production of protein arrays (5,85). More recently, we have also shown that the trans-splicing activity of the naturally occurring *Ssp* DnaE split-intein can be photomodulated by introducing photolabile backbone protecting groups on the C-intein polypeptide (119). This opens the intriguing possibility for light-activated immobilization of proteins onto solid supports, which should allow rapid production of protein arrays by using available photolithographic techniques (120).

PROTEIN ARRAY TECHNOLOGIES BASED ON CELL-FREE EXPRESSION SYSTEMS

Protein arrays have been traditionally produced by cellular expression, purification and immobilization of individual

proteins onto appropriate solid supports. The production of a large number of proteins using conventional expression systems, based on bacterial or eukaryotic cells, is usually a very time-consuming process that requires large amounts of manpower. Moreover, the presence of disulfide bonds, special requirements for folding and post-translational modifications in some proteins, especially those of human origin, may require more specialized expression systems such as mammalian cells or baculovirus. The stability of folded proteins in an immobilized state over long periods of storage is also another potential issue when working with protein microarrays, especially if we consider the highly heterogeneous nature of proteins in regards to their physicochemical properties and stability characteristics.

The use of cell-free expression systems has been proposed as a potential solution to circumvent some of these issues. Because DNA arrays can, in principle, be readily synthesized and are physically homogeneous and stable, the issues associated with availability and stability should not apply in this case. Hence, cell-free expression systems have the potential to allow the immobilization of proteins at the same time they are produced by converting DNA arrays into protein arrays on demand (7,121).

Cell-Free Protein Expression Systems

Cell-free expression systems make use of cell extracts that contain all of the key molecular components for carrying out transcription and translation *in vitro*. Typically, these extracts can be purified from cell lysates of different types of cells. The most commonly used are obtained from *E. coli*, rabbit reticulocyte and wheat germ, although more specialized cell extracts from hyperthermophiles, hybridomas, insect, and human cells can also be employed (7). This large variety of available cell-free expression systems ensures that proteins can be expressed under different conditions (122). Cell-free systems have also been used for the introduction of different biophysical probes during translation for protein detection and/or immobilization (123–125).

An important aspect to consider when preparing *in situ* protein arrays is the level of protein expression. While many proteins can be readily expressed, others may require modifications in the expression protocol or to the protein construct, for example by fusing them to a well-expressed fusion protein. He and co-workers have shown that using fusion protein constructs containing the constant domain of immunoglobulin κ light chain can significantly improve the expression levels of many proteins in *E. coli*-based cell-free expression systems (126).

Protein *In Situ* Array (PISA)

In this method, proteins are produced directly from DNA in solution and then immobilized as they are produced onto the surface through a recognition tag sequence (Fig. 7A) (127). In general, the DNA constructs encoding the proteins can be generated by PCR using designed specific primers for the protein of interest, although expression plasmids can also be used. The DNA constructs are also designed with strong promoters, such as T7, and regulatory sequences required for *in vitro* initiation of transcription/translation. An affinity tag sequence is also usually encoded into the N- or C-terminus of the protein to facilitate its immobilization after the translation step (Fig. 7A).

In this approach, all the proteins are expressed in parallel using the appropriate *in vitro* transcription/translation systems. The protein translation reaction is carried out on the surface, which is precoated with a capture reagent able to specifically bind to the affinity tag and immobilize the proteins. This is typically accomplished by using His-tagged proteins and Ni^{2+}-NTA coated surfaces, although other affinity tag/capture reagent combinations can also be used. Once the protein is translated and specifically immobilized onto the surface, any unbound material can be washed away.

The PISA method was originally demonstrated using a small set of proteins, which included several antibody fragments and the protein luciferase. These proteins were immobilized onto microliter wells and magnetic beads (127). In this work, PISA was used in a macro format in which ≈25 μL of cell-free expression reaction was used for the immobilization of individual proteins. More recently, PISA has also been miniaturized (using ≈40 nL) and adapted for the direct production of microarrays onto glass slides. In this method, the transcription/translation reaction is performed for 2 h at 30°C before spotting (7).

Hoheisel and co-workers have further developed the miniaturization of PISA using an on-chip system based on a multiple spotting technique (MIST) (128). In this approach, the DNA template is first spotted (≈350 pL) on the surface followed by the *in vitro* transcription/translation mixture on the same spot. The authors used His-tagged GFP as a model protein that was immobilized onto Ni^{2+}-NTA-coated glass slides. It was estimated that with unpurified PCR products, as little as 35 fg (≈22,500 molecules) of DNA was sufficient for the detection of GFP expression in sub-nL volumes (128). The same authors also adapted the system for the high throughput expression of libraries by designing a single specific primer pair for the introduction of the required T7 promoter and terminator, and demonstrated the *in situ* expression using 384 randomly chosen clones from a human fetal brain library (128). In principle, the optimized and miniaturized version of PISA should be able to produce high-density protein microarrays containing as much as 13,000 spots per slide using a variety of different genomic sources in a relatively uncomplicated fashion.

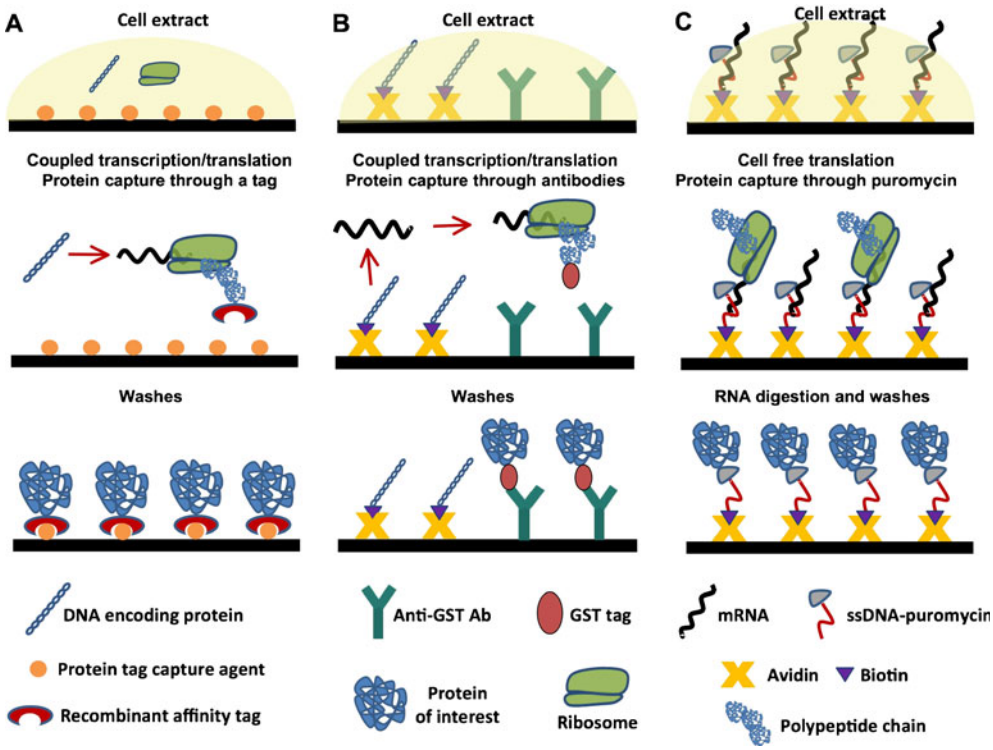


Fig. 7 *In situ* methods for protein arraying by PISA (**A**), NAPPA (**B**) and puromycin-capture from RNA arrays (**C**).

Nucleic Acid Programmable Protein Array (NAPPA)

NAPPA is another approach that allows the on-chip transformation of DNA arrays into protein arrays (Fig. 7B). NAPPA was initially developed by LaBaer and co-workers, and uses transcription and translation from an immobilized DNA template (67,129), as opposed to PISA, where the DNA template is kept in solution. In NAPPA, the expression plasmids encoding the proteins as GST fusions are biotinylated and immobilized onto a glass slide previously coated with avidin and an anti-GST antibody, which acts as the protein capture reagent. This plasmid array is then used for *in situ* expression of the proteins using rabbit reticulocyte lysate or a similar cell-free expression system. Once the proteins are translated, they are immediately captured by the immobilized antibody within each spot. This process generates a protein array in which every protein is co-localized with the corresponding expression plasmid. In general, NAPPA provides good quality protein spots with limited lateral spreading, although some variation can be observed in the quality of the arrays generated by this approach.

The first demonstration of the NAPPA approach was carried out by the immobilization of 8 different cell cycle proteins, which were immobilized at a density of 512 spots per slide (67). It was estimated that ≈10 fmol of protein were captured on average per spot, ranging from 4 to 29 fmol for the different proteins, which was sufficient for functional studies. The authors used this protein array to map and identify new interactions between 29 human proteins involved in initiation of DNA replication. These data were used to establish the regulation of Cdt1 binding to select replication proteins and map its geminin-binding domain (67).

As with PISA, NAPPA allows the protein array to be generated *in situ*, thus removing any concerns about protein stability during storage. However, it requires the cloning of the genes of interest and biotinylation of the resulting expression plasmids to facilitate their immobilization onto the chip (Fig. 7B). Furthermore, the technology does not generate a pure protein microarray, but rather a mixed array in which the different GST fusion proteins are co-localized with their corresponding expression plasmids, avidin and the capture antibody.

In Situ Puromycin-Capture from mRNA Arrays

Tao and Zhu have ingeniously adapted the mRNA display technology for the production protein of microarrays by capturing the nascent polypeptides through puromycin (Fig. 7C) (130). In this approach, the PCR-amplified DNA construct is transcribed into mRNA *in vitro*, and the 3′-end of the mRNA is hybridized with a single-stranded DNA oligonucleotide modified with biotin and puromycin. These modified RNAs are then arrayed on a streptavidin-coated glass slide and allowed to react with a cell-free lysate for *in vitro* translation. During the translation step, the ribosome stalls when it reaches the RNA/DNA hybrid section of the

molecule, and the DNA is then cross-linked to the nascent polypeptide through the puromycin moiety. Once the translation reaction is finished, the mRNA is digested with RNase, leaving a protein array immobilized through the C-termini to the DNA linker, which is immobilized through a biotin/streptavidin interaction to the surface. This technology was first exemplified by the immobilization of GST, two kinases, and two transcription factors (130). The transcription factors retained the ability to specifically bind DNA on the chip. This approach provides well-defined non-diffused protein spots as a result of the precise co-localization of the mRNA with puromycin and the 1:1 stoichiometry of mRNA *versus* protein. However, this method requires extra manipulations involving the reverse transcription and modification of the RNA before the spotting process, which may limit its practical use for the creation of large protein microarrays. Furthermore, the amount of protein produced is proportional to the amount of mRNA spotted, since there is no enzymatic amplification involved as in the PISA and NAPPA approaches.

DETECTION METHODS

In order to analyze, identify and quantify the proteins or any other type of biomolecules captured by the protein microarray, it is necessary to have detection methods that can provide high throughput analysis, high signal-to-noise ratio, good resolution, high dynamic range and reproducible results, with relatively low instrumentation costs. Most of the methods available for this task can be classified as label-dependent and label-free detection methods (see references (1,131) for recent reviews).

Label-Dependent Methods of Detection

Fluorescence-based detection is probably the most commonly used method in protein microarrays. This is mainly due to its simplicity, relatively high sensitivity and compatibility with already available DNA-array scanners. Protein-detecting microarrays usually employ a sandwich assay fluorescence-based detection system in which captured proteins are detected by a secondary fluorescent-labeled antibody (Fig. 1). This assay provides a higher specificity than the immunoassay based on a single antibody, since it reduces potential cross-reactivity issues. The sensitivity of fluorescence detection can also be improved by using the rolling circle amplification (RCA) method, which has been successfully applied for the profiling of different cytokines with detection limits on the fM range (132,133). The main limitation of these methods, however, is that they require two distinct capture reagents per protein to be analyzed, which means that if there are 1,000 proteins to be analyzed, more than 2,000 antibodies are required.

Specific fluorescence biosensing probes have also been used for the quantitative analysis of protein phosphorylation and protein kinase activity on functional protein microarrays. For example, the Pro-Q Diamond dye is a novel fluorescent phosphorylation sensor that allows the detection of phosphoproteins at sub-picogram levels of sensitivity (134). Hamachi and co-workers have also developed a fluorescence-based method for imaging monophosphorylated polypeptides by using bis-(Zn^{2+}-dipicolylamine)-based artificial sensors (135). Such chemical approaches do not require the use of anti-target antibodies and therefore represent a good approach for high throughput screening of protein phosphorylation and kinase activity.

The use of fluorescent-labeled substrates immobilized onto a microarray format has also been reported to study enzymatic specificity in a high throughput format. Ellman and co-workers have used this approach to determine the P-site substrate specificity of several serine and cysteine proteases (136). In their work, the fluorophore 7-amino-4-methy-coumarin (AMC) was covalently attached to a peptide microarray containing different amino acids at the different P-site positions. The corresponding sequence preferences were determined by analyzing the remaining fluorescence on the chip after performing the proteolytic reaction. Yao and co-workers have also used a similar approach for screening the activities of different types of enzymes, including proteases, epoxide hydrolases, and phosphatases by linking the substrate to the surface through a fluorogenic linker (137). The same authors have also developed a different approach for the activity-based detection of enzymes using a microarray format, in which the samples containing the enzymes to be analyzed are immobilized onto surfaces and then visualized with fluorescently labeled mechanism-based inhibitors (138).

The protein fingerprinting (PFP) technique is another fluorescence-based detection method that has been employed for the analysis of protein microarrays. This approach makes use of fluorophore-labeled capture reagents that change their fluorescent properties once they bind to the target protein; thus, by comparing patterns, the proteins of interest can be identified and at the same time discriminate any signal coming from non-specific interactions (139,140). This approach does not use high affinity capture reagents, such as antibodies, but rather uses relatively weak binders such as synthetic polypeptides.

Other label-dependent methods include the use of radioactivity, especially for enzymatic reactions such as phosphorylation, due to their sensitivity and specificity. For example, Schreiber and co-workers have used it to monitor kinase activity in combination with radioisotope-labeled

ATP (10). Snyder and co-workers have employed this approach to study the activities and substrate preferences of 119 different protein kinases (87). The use of radioisotope-labeled molecules, however, may raise safety concerns, thus limiting its potential for high throughput analysis. The use of chemilumiscence-based detection schemes also provides high selectivity and sensitivity, but with a limited resolution and dynamic range (141).

Label-Free Detection

The use of fluorescence-based detection methods is by far one of the most commonly employed approaches for the detection of proteins. However, there are several limitations to this approach. For example, labeling of proteins in samples or specific protein capture reagents such as antibodies may alter the surface of the proteins and therefore their binding properties. It is also a very time-consuming technique, especially when a multitude of samples need to be labeled. Another potential issue is the variability in labeling efficiency of proteins across different samples. This is a critical issue, especially when non-specific labeling techniques are employed, since small variations in the temperature and reaction duration, for example, can seriously influence the efficiency of protein labeling.

These limitations have sparked the development of novel label-free detection schemes involving mass spectrometry (MS)- and optical spectroscopy-based measurements (131,142).

In particular MS-based detection has already been used for the discovery of disease-associated biomarkers (143). For example, the use of surface-enhanced laser desorption ionization time-of-flight (SELDI-TOF) MS allows the detection of captured proteins without the need for labeling (144). In fact, SELDI has been widely used for the discovery and detection of biomarkers associated to several types of cancer (145–150). More recently, Becker and Engelhard have also used matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) for the direct read-out of protein/protein interactions using protein-DNA microarrays generated by DNA-directed immobilization (DDI) (61) (see reference (62) for a recent review in this field). The authors used this approach for the rapid detection of activated Ras in cell lysates from several cell lines.

Another well-established label-free detection method is surface plasmon resonance (SPR). SPR can also provide kinetic information on binding events. In this approach, the appropriate capture reagents are immobilized onto a gold surface, and quantification of the captured proteins is carried out by measuring the change in the reflection angle of light after hitting the gold surface (151). For example, an SPR imaging method was recently used for the high throughput screening of molecules able to target the interaction between the retinoblastoma tumor suppressor RB and the human papillomavirus (HPV) E7 proteins (152). The E7 protein is produced by high-risk human papillomavirus (HPV) and induces degradation of the retinoblastoma tumor suppressor RB through a direct interaction, and it has been suggested as a potential molecular target in cancer therapy. In this work, a glutathione-coated SPR chip was used for the immobilization of the E7 GST-fusion protein, which was then complexed with His-tagged RB protein in the presence of different RB-binding peptides derived from a motif of the E7 protein. Some of these peptides were shown to antagonize the interaction between His-tagged RB and GST-E7 in a concentration-dependent manner (152).

A conventional SPR system, however, can only use a single channel per experiment. The recent development of SPR microscopy allows the analysis of hundreds of biomolecular interactions simultaneously in large protein microarrays (>1,300 spots) allowing for qualitative screening and quantitative kinetics experiments in a high throughput format (153).

The anomalous reflection (AR) technique is another spectroscopic detection scheme that has been suggested as an alternative to SPR. AR is a characteristic property of gold that causes a large decrease in the reflectivity of blue or purple light ($380\ \text{nm} < \lambda < 480\ \text{nm}$) on a gold surface upon adsorption of a transparent dielectric layer on its surface (154). The AR technique requires relatively less complex optics than the SPR systems and has the potential to offer miniaturized and parallelized measurements; therefore, it could be potentially suitable as a high-throughput analytical platform. This approach has been used so far with some success for analyzing biotin/avidin, calmodulin/synthetic α-helical peptides and T7-phage displayed-proteins and synthetic peptide interactions (154,155). At this point, however, AR-based detection of microarrays needs to be further developed for detection of multiplexed protein-protein interactions beyond the proof-of-concept.

APPLICATIONS

Some of the applications of protein microarrays have already been discussed in the previous sections. The most prominent applications include high-throughput proteomics, biomarker research and drug discovery. Several reviews focusing on the biomedical applications of protein microarrays have been published recently (3,8).

Proteomics

Functional protein microarrays are ideal bioanalytical platforms to carry out high-throughput proteomics.

Perhaps the most advanced example of this application to date was reported by MacBeath and co-workers to study the phosphorylation states of the ErbB-receptor kinase family using functional protein microarrays (23,156). The first three members of the ErbB family of receptor tyrosine kinases, ErbB1-3, are involved in the activation of a wide variety of signaling pathways that are frequently misregulated in cancer. Erb4, on the other hand, is not involved in tumorigenesis and has been shown to have a protective role in some cancers. In order to study in more detail the role of this receptor tyrosine kinase, the authors first used tandem mass spectrometry to identify 19 sites of tyrosine phosphorylation on ErbB4. These phosphopeptides were then used to probe a functional protein microarray containing 96 SH2 and 37 PTB protein domains encoded in the human genome. The obtained data was used to build a quantitative interaction network for ErbB4 as well as for the identification of several new interactions that led to the finding that ErbB4 can bind and activate STAT1 (Fig. 2).

Deng and co-workers have also studied protein-protein and protein-DNA interactions on a global scale in the plant *A. thaliana* by making use of functional microarrays (157). The authors created a microarray containing up to 802 different transcription factors from *A. thaliana*. The proteins were expressed using a yeast expression system and arrayed onto FAST glass slides, which are commercially available slides coated with a nitrocellulose membrane. The resulting microarray was probed with different fluorescent-labeled oligonucleotides containing known binding sites for several transcription factors of the AP2/ERF family. Using this approach the authors were able to confirm known interactions and identify 48 new ones. These included four transcription factors that were able to bind the evening element and showed an expected clock-regulated gene expression pattern, thus providing a basis for further functional analysis of their roles in circadian-regulated gene expression (157). The same authors also used this microarray for detecting novel protein-protein interactions and were able to discover four novel partners that interact with transcription factor HY5 (157), which is a key regulator of photomorphogenesis in *A. thaliana* (157).

It should be highlighted, however, that the production of whole-proteome microarrays is technically a challenging task, since it requires the isolation of a large number of functional proteins. Furthermore, the analysis of whole-genome microarrays is complicated due to the fact that they only represent particular time snapshots of the proteome. Moreover, proteins not only differ in structure and function but also in their cellular localization, turnover rates and, more importantly, abundance. However, the use of this technology in proteomic research still allows the unprecedented ability to monitor the biomolecular interactions of thousands of samples in parallel, which by far outweighs all the difficulties and limitations associated with their use and preparation.

Biomarker Research

The use of protein microarrays in biomarker research has received special interest in the areas of viral diagnostics and cancer research. For example, the examination and identification of particular protein profiles in early-stage cancers could lead to early detection of tumors and the development of improved therapies for cancer patients. Antibody-based microarrays are by far the most frequently used in biomarker profiling and discovery for cancer research. For example, Cordon-Cardo and co-workers have used an antibody array composed by 254 different antibodies to discriminate bladder cancer patients from control patients (40). Snyder and co-workers have also used protein microarrays to profile antibodies against human severe acute respiratory syndrome (SARS) virus and related coronaviruses (158). In their study, the authors used 82 different coronavirus GST-fusion proteins, which were expressed in yeast and arrayed onto FAST glass slides. These arrays were used to profile the sera of two patient groups (more than 600 samples obtained from patients in China and Canada) with ≈90% accuracy (158). Using this approach, it was possible to distinguish patients infected with SARS and HCoV-229E, two different human coronavirus. These results were further validated by statistical methods and an indirect immuno-fluorescence test, and also showed that the sensitivity provided by microarray profiling was similar in sensitivity to standard indirect immuno-fluorescence tests but was more specific (158).

LaBaer and co-workers have also used protein microrrays generated by the NAPPA approach for tumor antigen profiling in breast cancer (16). In this work, sera from breast and ovarian cancer patients were tested for p53-specific antibodies using a microarray displaying 1,705 different non-redundant tumor antigens. These results were also corroborated by standard indirect immuno-blotting techniques (16).

The described examples are just a sample of the recent applications of protein microarrays in biomarker profiling and discovery, and illustrate the great potential of this technology in biomedical applications.

Drug Discovery

Protein microarrays have also been used in drug discovery for target identification and validation. In 2004, Schreiber and co-workers described for the first time the use of a protein microarray for high-throughput screening of small molecules (159). In this work, the authors used a protein microarray obtained by spotting different His-tagged and GST-fusion proteins onto chemically modified glass slides.

These arrays were used to screen the molecular targets of six small-molecule inhibitors of rapamycin (SMIR) that were previously identified for their ability to rescue growth of yeast cells exposed to rapamycin in a phenotype-based chemical genetic suppressor assay. To facilitate the screening process, the SMIRs were conjugated to biotin, and the bound SMIRs were then detected using fluorescent-labeled streptavidin. These results allowed the identification of a new, unknown member of the target of rapamycin (TOR) signaling pathway (159).

Protein microarrays can also be used in an indirect fashion for screening and selecting small molecules able to antagonize protein interactions. For example, antibody arrays can be used to screen and/or profile the proteome for changes in protein expression and/or post-translational modifications, such as phosphorylation, induced by the presence or absence of a particular drug candidate.

Sokolov and Cadet have used protein microarrays to study the correlation between the levels of expression of different proteins and the behavioral phenotype of mice treated with methamphetamine (METH) (160). METH abuse has been shown to stimulate aggressive behaviors in humans and in other animals. The authors found that mice treated chronically with METH demonstrated increased aggressiveness and hyper-locomotion when compared to an untreated control group. In this work, a total of 378 different monoclonal antibodies specific for proteins related to signal transduction, oncogene products, cell cycle regulation, cell structure, apoptosis, and neurobiology, among others, were used to prepare the protein-detecting array (160). This antibody microarray was incubated with proteins extracted from the brain of untreated and METH-treated mice and labeled with fluorescent dyes. The data revealed a decrease in the natural abundance of the proteins Erk2 and 14-3-3e in the striata of the mice chronically treated with METH. Since the kinase Erk2 is thought to be the principal component of the classical mithogen-activated protein (MAP) kinase pathway and protein 14-3-3e is an inhibitor and substrate of protein kinase C, the reduction in these two proteins suggests that repeated exposure to METH might alter MAP kinase-related pathways involved in behavioral change (160).

These examples clearly illustrate the potential of protein microarrays for drug discovery applications. Despite the numerous advantages in the preparation and analysis of these types of reagents, their use in drug discovery has been limited so far.

CONCLUDING REMARKS

The aim of this review is to highlight the latest developments in the preparation, analysis and biotechnological applications of protein microarrays. Just before MacBeath and Schreiber reported for the first time the use of protein microarrays in 2000 (10), the concept of using protein microarray technology was simply regarded as a dream. A decade later, the number of publications on protein microarray technologies has increased dramatically. There are approximately 32,000 publications indexed in PubMed (http://www.ncbi.nlm.nih.gov/pubmed) under the keyword *protein microarrays*. We have seen numerous examples that show protein microarrays are a very valuable tool for the study of whole proteomes (11–13,18,23,24), protein identification and profiling for early diagnosis of diseases such as cancer (16,40) or viral infections (158) and for drug identification and validation (159,160).

Despite the large number of successful examples in the use of protein microarrays in biomedical and biotechnological applications during the last 10 years, there are still, however, some challenges that need to be tackled. For example, most of the methods commonly employed for the immobilization of proteins onto solid supports rely on non-site-specific immobilization techniques (10,46,47,49,161). The use of these methods usually results in the proteins being displayed in random orientations on the surface, which may compromise the biological activity of the immobilized proteins and/or provide false results (162). This issue has been addressed over the last few years by the development of novel site-specific immobilization approaches which involve the use of chemoselective ligation reactions (52,84,92,97,101,102), active site-directed capture ligands (112,116,163–165) and protein splicing (68,85), among others.

The expression and purification of thousands of proteins without compromising their structural and biological activity is also a challenging task. The use of cell-free expression systems in combination with nucleic acid arrays, which are more readily available and easier to prepare, has been shown to give good results to produce *in situ* protein arrays from DNA (67,127,129) and RNA arrays (130). The combination of these approaches with site-specific and traceless methods of protein immobilization such as protein trans-splicing (68,85) shows great promise.

The introduction of label-free detection methods, such as surface plasmon resonance and mass spectrometry, also shows great promise to simplify the use of protein microarray analysis, since labeling of the interacting partners will no longer be required.

The standardization of protein microarray production is another issue that needs to be improved. At this time, most of the methods used by the scientific community for preparing and analyzing protein microarrays are not completely standardized. The adoption of stringent standards by the scientific community for the production and analysis of these valuable reagents should, in principle,

allow the generation of data that could be compared and exchanged across different studies and different research groups.

None of these challenges is impossible to achieve; in fact, as we have seen in this review, much more progress has been made over the last decade to address them. At this point, we strongly believe that the protein microarray technology is on the brink of becoming a standard technique in research in the same way as DNA microarray technology is used today.

ACKNOWLEDGEMENTS

Work was supported by funding from the School of Pharmacy at the University of Southern California, National Institute of Health award GM090323-01 and Department of Defense, Prostate Cancer Research Program of the Office of the Congressionally Directed Medical Research Programs award PC093051 to J.A.C.

REFERENCES

1. Tomizaki KY, Usui K, Mihara H. Protein-detecting microarrays: current accomplishments and requirements. Chembiochem. 2005;6(5):782–99.
2. Wolf-Yadlin A, Sevecka M, MacBeath G. Dissecting protein function and signaling using protein microarrays. Curr Opin Chem Biol. 2009;13(4):398–405.
3. Weinrich D, Jonkheijm P, Niemeyer CM, Waldmann H. Applications of protein biochips in biomedical and biotechnological research. Angew Chem Int Ed Engl. 2009;48(42):7744–51.
4. Camarero JA. Recent developments in the site-specific immobilization of proteins onto solid supports. Biopolymers. 2008;90 (3):450–8.
5. Coleman MA, Hoeprich P, Beernink P, Camarero JA. Cell-free protein expression screening and protein immobilization using protein microarrays. In: Kudlicki W, Katzen F, Bennett R, editors. Cell Free Expression Systems Landes Bioscience Publishers; 2007.
6. Coleman MA, Beernink PT, Camarero JA, Albala JS. Applications of functional protein microarrays: identifying protein–protein interactions in an array format. Methods Mol Biol. 2007;385:121–30.
7. He M, Stoevesandt O, Taussig MJ. *In situ* synthesis of protein arrays. Curr Opin Biotechnol. 2008;19(1):4–9.
8. Spisak S, Guttman A. Biomedical applications of protein microarrays. Curr Med Chem. 2009;16(22):2806–15.
9. Kodadek T. Protein microarrays: prospects and problems. Chem Biol. 2001;8(2):105–15.
10. MacBeath G, Schreiber SL. Printing proteins as microarrays for high-throughput function determination. Science. 2000;289 (5485):1760–3.
11. Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, *et al.* Global analysis of protein activities using proteome chips. Science. 2001;293(5537):2101–5.
12. Ptacek J, Devgan G, Michaud G, Zhu H, Zhu X, Fasolo J, *et al.* Global analysis of protein phosphorylation in yeast. Nature. 2005;438(7068):679–84.
13. Popescu SC, Popescu GV, Bachan S, Zhang Z, Gerstein M, Snyder M, *et al.* MAPK target networks in Arabidopsis thaliana revealed using functional protein microarrays. Genes Dev. 2009;23(1):80–92.
14. Lin YY, Lu JY, Zhang J, Walter W, Dang W, Wan J, *et al.* Protein acetylation microarray reveals that NuA4 controls key metabolic target regulating gluconeogenesis. Cell. 2009;136 (6):1073–84.
15. Hudson ME, Pozdnyakova I, Haines K, Mor G, Snyder M. Identification of differentially expressed proteins in ovarian cancer using high-density protein microarrays. Proc Natl Acad Sci USA. 2007;104(44):17494–9.
16. Anderson KS, Ramachandran N, Wong J, Raphael JV, Hainsworth E, Demirkan G, *et al.* Application of protein microarrays for multiplexed detection of antibodies to tumor antigens in breast cancer. J Proteome Res. 2008;7(4):1490–9.
17. Merbl Y, Kirschner MW. Large-scale detection of ubiquitination substrates using cell extracts and protein microarrays. Proc Natl Acad Sci USA. 2009;106(8):2543–8.
18. Espejo A, Cote J, Bednarek A, Richard S, Bedford MT. A protein-domain microarray identifies novel protein–protein interactions. Biochem J. 2002;367(Pt 3):697–702.
19. Hesselberth JR, Miller JP, Golob A, Stajich JE, Michaud GA, Fields S. Comparative analysis of Saccharomyces cerevisiae WW domains and their interacting proteins. Genome Biol. 2006;7(4):R30.
20. Lim CS, Seet BT, Ingham RJ, Gish G, Matskova L, Winberg G, *et al.* The K15 protein of Kaposi's sarcoma-associated herpesvirus recruits the endocytic regulator intersectin 2 through a selective SH3 domain interaction. Biochemistry. 2007;46 (35):9874–85.
21. Polverini E, Rangaraj G, Libich DS, Boggs JM, Harauz G. Binding of the proline-rich segment of myelin basic protein to SH3 domains: spectroscopic, microarray, and modeling studies of ligand conformation and effects of posttranslational modifications. Biochemistry. 2008;47(1):267–82.
22. Boutell JM, Hart DJ, Godber BL, Kozlowski RZ, Blackburn JM. Functional protein microarrays for parallel characterisation of p53 mutants. Proteomics. 2004;4(7):1950–8.
23. Jones RB, Gordus A, Krall JA, MacBeath G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature. 2006;439(7073):168–74.
24. Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, Allen JE, *et al.* PDZ domain binding selectivity is optimized across the mouse proteome. Science. 2007;317(5836):364–9.
25. Chen JR, Chang BH, Allen JE, Stiffler MA, MacBeath G. Predicting PDZ domain-peptide interactions from primary sequences. Nat Biotechnol. 2008;26(9):1041–5.
26. Kaushansky A, Allen JE, Gordus A, Stiffler MA, Karp ES, Chang BH, *et al.* Quantifying protein–protein interactions in high throughput using protein domain microarrays. Nat Protoc. 2010;5(4):773–90.
27. Woo Y-H, Camarero JA. Interfacing 'Hard' and 'Soft' matter with exquisite chemical control. Curr Nanosci. 2006;2:93–103.
28. Camarero JA. New developments for the site-specific attachment of proptein to surface. Biophys Rev Lett. 2006;1(1):1–28.
29. Jonkheijm P, Weinrich D, Schroder H, Niemeyer CM, Waldmann H. Chemical strategies for generating protein biochips. Angew Chem Int Ed. 2008;47(50):9618–47.
30. Lin PC, Weinrich D, Waldmann H. Protein biochips: oriented surface immobilization of proteins. Macromol Chem Phys. 2010;211(2):136–44.
31. Kolmar H, Skerra A. Alternative binding proteins get mature: rivalling antibodies. FEBS J. 2008;275(11):2667.
32. Daly NL, Rosengren KJ, Craik DJ. Discovery, structure and biological activities of cyclotides. Adv Drug Deliv Rev. 2009;61 (11):918–30.

33. Miller JC, Zhou H, Kwekel J, Cavallo R, Burke J, Butler EB, *et al.* Antibody microarray profiling of human prostate cancer sera: antibody screening and identification of potential biomarkers. Proteomics. 2003;3(1):56–63.
34. Gembitsky DS, Lawlor K, Jacovina A, Yaneva M, Tempst P. A prototype antibody microarray platform to monitor changes in protein tyrosine phosphorylation. Mol Cell Proteomics. 2004;3 (11):1102–18.
35. Nielsen UB, Cardone MH, Sinskey AJ, MacBeath G, Sorger PK. Profiling receptor tyrosine kinase activation by using Ab microarrays. Proc Natl Acad Sci USA. 2003;100(16):9330–5.
36. Chen S, LaRoche T, Hamelinck D, Bergsma D, Brenner D, Simeone D, *et al.* Multiplexed analysis of glycan variation on native proteins captured by antibody microarrays. Nat Methods. 2007;4(5):437–44.
37. Kusnezow W, Banzon V, Schroder C, Schaal R, Hoheisel JD, Ruffer S, *et al.* Antibody microarray-based profiling of complex specimens: systematic evaluation of labeling strategies. Proteomics. 2007;7(11):1786–99.
38. Ghobrial IM, McCormick DJ, Kaufmann SH, Leontovich AA, Loegering DA, Dai NT, *et al.* Proteomic analysis of mantle-cell lymphoma by protein microarray. Blood. 2005;105 (9):3722–30.
39. Hamelinck D, Zhou H, Li L, Verweij C, Dillon D, Feng Z, *et al.* Optimized normalization for antibody microarrays and application to serum-protein profiling. Mol Cell Proteomics. 2005;4 (6):773–84.
40. Sanchez-Carbayo M, Socci ND, Lozano JJ, Haab BB, Cordon-Cardo C. Profiling bladder cancer using targeted antibody arrays. Am J Pathol. 2006;168(1):93–103.
41. Gaudet S, Janes KA, Albeck JG, Pace EA, Lauffenburger DA, Sorger PK. A compendium of signals and responses triggered by prodeath and prosurvival cytokines. Mol Cell Proteomics. 2005;4 (10):1569–90.
42. Janes KA, Albeck JG, Gaudet S, Sorger PK, Lauffenburger DA, Yaffe MB. A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. Science. 2005;310 (5754):1646–53.
43. Paweletz CP, Charboneau L, Bichsel VE, Simone NL, Chen T, Gillespie JW, *et al.* Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. Oncogene. 2001;20 (16):1981–9.
44. Chan SM, Ermann J, Su L, Fathman CG, Utz PJ. Protein microarrays for multiplex analysis of signal transduction pathways. Nat Med. 2004;10(12):1390–6.
45. Ambroz KL, Zhang Y, Schutz-Geschwender A, Olive DM. Blocking and detection chemistries affect antibody performance on reverse phase protein arrays. Proteomics. 2008;8(12):2379–83.
46. Arenkov P, Kukhtin A, Gemmell A, Voloshchuk S, Chupeeva V, Mirzabekov A. Protein microchips: use for immunoassay and enzymatic reactions. Anal Biochem. 2000;278(2):123–31.
47. Lee KB, Park SJ, Mirkin CA, Smith JC, Mrksich M. Protein nanoarrays generated by dip-pen nanolithography. Science. 2002;295(5560):1702–5.
48. Chan WC, Nie S. Quantum dot bioconjugates for ultrasensitive nonisotopic detection. Science. 1998;281(5385):2016–8.
49. Liu GY, Amro NA. Positioning protein molecules on surfaces: a nanoengineering approach to supramolecular chemistry. Proc Natl Acad Sci USA. 2002;99(8):5165–70.
50. Phizicky E, Bastiaens PI, Zhu H, Snyder M, Fields S. Protein analysis on a proteomic scale. Nature. 2003;422(6928):208–15.
51. Lesaicherre ML, Lue RY, Chen GY, Zhu Q, Yao SQ. Intein-mediated biotinylation of proteins and its application in a protein microarray. J Am Chem Soc. 2002;124(30):8768–9.
52. Lesaicherre ML, Uttamchandani M, Chen GY, Yao SQ. Developing site-specific immobilization strategies of peptides in a microarray. Bioorg Med Chem Lett. 2002;12(16):2079–83.
53. Delehanty JB, Ligler FS. A microarray immunoassay for simultaneous detection of proteins and bacteria. Anal Chem. 2002;74(21):5681–7.
54. Marin VL, Bayburt TH, Sligar SG, Mrksich M. Functional assays of membrane-bound proteins with SAMDI-TOF mass spectrometry. Angew Chem Int Ed. 2007;46(46):8796–8.
55. Kawahashi Y, Doi N, Takashima H, Tsuda C, Oishi Y, Oyama R, *et al. In vitro* protein microarrays for detecting protein–protein interactions: application of a new method for fluorescence labeling of proteins. Proteomics. 2003;3(7):1236–43.
56. Quinn J, Patel P, Fitzpatrick B, Manning B, Dillon P, Daly S, *et al.* The use of regenerable, affinity ligand-based surfaces for immunosensor applications. Biosens Bioelectron. 1999;14 (6):587–95.
57. Oh BK, Kim YK, Lee W, Bae YM, Lee WH, Choi JW. Immunosensor for detection of Legionella pneumophila using surface plasmon resonance. Biosens Bioelectron. 2003;18(5–6):605–11.
58. Dickason RR, Edwards RA, Bryan J, Huston DP. Versatile E. coli thioredoxin specific monoclonal antibodies afford convenient analysis and purification of prokaryote expressed soluble fusion protein. J Immunol Methods. 1995;185(2):237–44.
59. di Guan C, Li P, Riggs PD, Inouye H. Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. Gene. 1988;67(1):21–30.
60. Chong S, Mersha FB, Comb DG, Scott ME, Landry D, Vence LM, *et al.* Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. Gene. 1997;192(2):271–81.
61. Becker CF, Wacker R, Bouschen W, Seidel R, Kolaric B, Lang P, *et al.* Direct readout of protein–protein interactions by mass spectrometry from protein–DNA microarrays. Angew Chem Int Ed Engl. 2005;44(46):7635–9.
62. Niemeyer CM. Semisynthetic DNA–protein conjugates for biosensing and nanofabrication. Angew Chem Int Ed Engl. 2010;49(7):1200–16.
63. Soellner MB, Dickson KA, Nilsson BL, Raines RT. Site-specific protein immobilization by Staudinger ligation. J Am Chem Soc. 2003;125(39):11790–1.
64. Kohn M, Wacker R, Peters C, Schroder H, Soulere L, Breinbauer R, *et al.* Staudinger ligation: a new immobilization strategy for the preparation of small-molecule arrays. Angew Chem Int Ed. 2003;42(47):5830–4.
65. Lue RY, Chen GY, Hu Y, Zhu Q, Yao SQ. Versatile protein biotinylation strategies for potential high-throughput proteomics. J Am Chem Soc. 2004;126(4):1055–62.
66. Yin J, Liu F, Li X, Walsh CT. Labeling proteins with small molecules by site-specific posttranslational modification. J Am Chem Soc. 2004;126(25):7754–5.
67. Ramachandran N, Hainsworth E, Bhullar B, Eisenstein S, Rosen B, Lau AY, *et al.* Self-assembling protein microarrays. Science. 2004;305(5680):86–90.
68. Camarero JA, Kwon Y. Traceless and site-specific attachment of proteins onto solid supports. Int J Pept Res Ther. 2008;14:351–7.
69. Lin P-C, Weinrich D, Waldmann H. Protein biochips: oriented surface immobilization of proteins. Macromol Chem Phys. 2010;211:136–44.
70. Schnolzer M, Kent SB. Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease. Science. 1992;256(5054):221–5.
71. Kohn M, Breinbauer R. The staudinger ligation—a gift to chemical biology. Angew Chem Int Ed. 2004;43(24):3106–16.

72. Muir TW. Semisynthesis of proteins by expressed protein ligation. Annu Rev Biochem. 2003;72:249–89.
73. Tam JP, Xu JX, Eom KD. Methods and strategies of peptide ligation. Biopolymers. 2001;60(3):194–205.
74. Aimoto S. Contemporary methods for peptide and protein synthesis. Curr Org Chem. 2001;5(1):45–87.
75. Nilsson BL, Kiessling LL, Raines RT. High-yielding Staudinger ligation of a phosphinothioester and azide to form a peptide. Org Lett. 2001;3(1):9–12.
76. Rostovtsev VV, Green LG, Fokin VV, Sharpless KB. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. Angew Chem Int Ed. 2002;41(14):2596–9.
77. Wang Q, Chan TR, Hilgraf R, Fokin VV, Sharpless KB, Finn MG. Bioconjugation by copper(I)-catalyzed azide-alkyne [3+2] cycloaddition. J Am Chem Soc. 2003;125(11):3192–3.
78. Rusmini F, Zhong Z, Feijen J. Protein immobilization strategies for protein biochips. Biomacromolecules. 2007;8(6):1775–89.
79. Heise A, Menzel H, Yim H, Foster MD, Wieringa RH, Schouten AJ, *et al.* Grafting of polypeptides on solid substrates by initiation of N-carboxyanhydride polymerization by amino-terminated self-assembled monolayers. Langmuir. 1997;13 (4):723–8.
80. Ulman A. Formation and structure of self-assembled monolayers. Chem Rev. 1996;96:1533–54.
81. Sullivan TP, Huck WTS. Reactions on monolayers: organic synthesis in two dimensions. Eur J Org Chem. 2003;2003:17–29.
82. Biebuyck HA, Bian CD, Whitesides GM. Comparison of organic monolayers on plycrystalline gold spontaneously. Langmuir. 1994;10(6):1825–31.
83. Cheung CL, Camarero JA, Woods BW, Lin TW, Johnson JE, De Yoreo JJ. Fabrication of assembled virus nanostructures on templates of chemoselective linkers formed by scanning probe nanolithography. J Am Chem Soc. 2003;125(23):6848–9.
84. Camarero JA, Kwon Y, Coleman MA. Chemoselective attachment of biologically active proteins to surfaces by expressed protein ligation and its application for "protein chip" fabrication. J Am Chem Soc. 2004;126(45):14730–1.
85. Kwon Y, Coleman MA, Camarero JA. Selective immobilization of proteins onto solid supports through split-intein-mediated protein trans-splicing. Angew Chem Int Ed. 2006;45(11):1726–9.
86. Prats-Alfonso E, Garcia-Martin F, Bayo N, Cruz LJ, Pla-Roca M, Samitier J, *et al.* Facile solid-phase synthesis of biotinylated alky thiols. Tetrahedron. 2006;62(29):6876–81.
87. Zhu H, Klemic JF, Chang S, Bertone P, Casamayor A, Klemic KG, *et al.* Analysis of yeast protein kinases using protein chips. Nat Genet. 2000;26(3):283–9.
88. Delamarche E, Bernard A, Schmid H, Michel B, Biebuyck H. Patterned delivery of immunoglobulins to surfaces using microfluidic networks. Science. 1997;276(5313):779–81.
89. Nishikawa M, Yamamoto T, Kojima N, Kikuo K, Fujii T, Sakai Y. Stable immobilization of rat hepatocytes as hemispheroids onto collagen-conjugated poly-dimethylsiloxane (PDMS) surfaces: importance of direct oxygenation through PDMS for both formation and function. Biotechnol Bioeng. 2008;99(6):1472–81.
90. Henry AC, Tutt TJ, Galloway M, Davidson YY, McWhorter CS, Soper SA, *et al.* Surface modification of poly(methyl methacrylate) used in the fabrication of microanalytical devices. Anal Chem. 2000;72(21):5331–7.
91. Noren CJ, Wang JM, Perler FB. Dissecting the chemistry of protein splicing and its applications. Angew Chem Int Ed. 2000;39(3):451–6.
92. Girish A, Sun H, Yeo DS, Chen GY, Chua TK, Yao SQ. Site-specific immobilization of proteins in a microarray using intein-mediated protein splicing. Bioorg Med Chem Lett. 2005;15 (10):2447–51.
93. Frank R, Overwin H. SPOT synthesis. Epitope analysis with arrays of synthetic peptides prepared on cellulose membranes. Methods Mol Biol. 1996;66:149–69.
94. Toepert F, Knaute T, Guffler S, Pires JR, Matzdorf T, Oschkinat H, *et al.* Combining SPOT synthesis and native peptide ligation to create large arrays of WW protein domains. Angew Chem Int Ed. 2003;42(10):1136–40.
95. Saxon E, Armstrong JI, Bertozzi CR. A "traceless" Staudinger ligation of the chemoselective synthesis of amide bonds. Org Lett. 2000;2(14):2141–3.
96. Saxon E, Bertozzi CR. Cell surface engineering by a modified Staudinger reaction. Science. 2000;287(5460):2007–10.
97. Watzke A, Kohn M, Gutierrez-Rodriguez M, Wacker R, Schroder H, Breinbauer R, *et al.* Site-selective protein immobilization by staudinger ligation. Angew Chem Int Ed. 2006;45 (9):1408–12.
98. Kiick KL, Saxon E, Tirrell DA, Bertozzi CR. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. Proc Natl Acad Sci USA. 2002;99 (1):19–24.
99. Link AJ, Vink MK, Tirrell DA. Presentation and detection of azide functionality in bacterial cell surface proteins. J Am Chem Soc. 2004;126(34):10598–602.
100. Kalia J, Raines RT. Reactivity of intein thioesters: appending a functional group to a protein. Chembiochem. 2006;7(9):1375–83.
101. Lin PC, Ueng SH, Tseng MC, Ko JL, Huang KT, Yu SC, *et al.* Site-specific protein modification through Cu(I)-catalyzed 1, 2, 3-triazole formation and its implementation in protein microarray fabrication. Angew Chem Int Ed. 2006;45 (26):4286–90.
102. Gauchet C, Labadie GR, Poulter CD. Regio- and chemoselective covalent immobilization of proteins through unnatural amino acids. J Am Chem Soc. 2006;128(29):9274–5.
103. Duckworth BP, Xu J, Taton TA, Guo A, Distefano MD. Site-specific, covalent attachment of proteins to a solid surface. Bioconjug Chem. 2006;17(4):967–74.
104. Taki M, Sisido M. Leucyl/phenylalanyl(L/F)-tRNA-protein transferase-mediated aminoacyl transfer of a nonnatural amino acid to the N-terminus of peptides and proteins and subsequent functionalization by bioorthogonal reactions. Biopolymers. 2007;88(2):263–71.
105. Watanabe K, Toh Y, Suto K, Shimizu Y, Oka N, Wada T, *et al.* Protein-based peptide-bond formation by aminoacyl-tRNA protein transferase. Nature. 2007;449(7164):867–71.
106. Ebisu K, Tateno H, Kuroiwa H, Kawakami K, Ikeuchi M, Hirabayashi J, *et al.* N-terminal specific point-immobilization of active proteins by the one-pot NEXT-A method. Chembiochem. 2009;10(15):2460–4.
107. Connor RE, Piatkov K, Varshavsky A, Tirrell DA. Enzymatic N-terminal addition of noncanonical amino acids to peptides and proteins. Chembiochem. 2008;9(3):366–9.
108. Codelli JA, Baskin JM, Agard NJ, Bertozzi CR. Second-generation difluorinated cyclooctynes for copper-free click chemistry. J Am Chem Soc. 2008;130(34):11486–93.
109. Baskin JM, Prescher JA, Laughlin ST, Agard NJ, Chang PV, Miller IA, *et al.* Copper-free click chemistry for dynamic *in vivo* imaging. Proc Natl Acad Sci USA. 2007;104(43):16793–7.
110. Ning X, Guo J, Wolfert MA, Boons GJ. Visualizing metabolically labeled glycoconjugates of living cells by copper-free and fast huisgen cycloadditions. Angew Chem Int Ed. 2008;47 (12):2253–5.
111. Govindaraju T, Jonkheijm P, Gogolin L, Schroeder H, Becker CF, Niemeyer CM, *et al.* Surface immobilization of biomolecules by click sulfonamide reaction. Chem Commun (Camb). 2008; (32):3723–5.

112. Hodneland CD, Lee YS, Min DH, Mrksich M. Supramolecular chemistry and self-assembly special feature: selective immobilization of proteins to self-assembled monolayers presenting active site-directed capture ligands. Proc Natl Acad Sci USA. 2002;99 (8):5048–52.
113. Mannesse ML, Boots JW, Dijkman R, Slotboom AJ, van der Hijden HT, Egmond MR, *et al.* Phosphonate analogues of triacylglycerols are potent inhibitors of lipase. Biochim Biophys Acta. 1995;1259(1):56–64.
114. Kwon Y, Han Z, Karatan E, Mrksich M, Kay BK. Antibody arrays prepared by cutinase-mediated immobilization on self-assembled monolayers. Anal Chem. 2004;76(19):5713–20.
115. Keppler A, Pick H, Arrivoli C, Vogel H, Johnsson K. Labeling of fusion proteins with synthetic fluorophores in live cells. Proc Natl Acad Sci USA. 2004;101(27):9955–9.
116. Sielaff I, Arnold A, Godin G, Tugulu S, Klok HA, Johnsson K. Protein function microarrays based on self-immobilizing and self-labeling fusion proteins. Chembiochem. 2006;7(1):194–202.
117. Lew BM, Mills KV, Paulus H. Protein splicing *in vitro* with a semisynthetic two-component minimal intein. J Biol Chem. 1998;273(26):15887–90.
118. Perler FB. A natural example of protein trans-splicing. Trends Biochem Sci. 1999;24(6):209–11.
119. Berrade L, Kwon Y, Camarero JA. Photomodulation of protein trans-splicing through backbone photocaging of the DnaE split intein. Chembiochem. 2010;cbic_201000157.
120. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. Science. 1991;251(4995):767–73.
121. He M, Stoevesandt O, Palmer EA, Khan F, Ericsson O, Taussig MJ. Printing protein arrays from DNA arrays. Nat Methods. 2008;5(2):175–7.
122. Endoh T, Kanai T, Sato YT, Liu DV, Yoshikawa K, Atomi H, *et al.* Cell-free protein synthesis at high temperatures using the lysate of a hyperthermophile. J Biotechnol. 2006;126(2):186–95.
123. Ozawa K, Wu PS, Dixon NE, Otting G. N-Labelled proteins by cell-free protein synthesis. Strategies for high-throughput NMR studies of proteins and protein-ligand complexes. FEBS J. 2006;273(18):4154–9.
124. Ohta A, Yamagishi Y, Suga H. Synthesis of biopolymers using genetic code reprogramming. Curr Opin Chem Biol. 2008;12 (2):159–67.
125. Goto Y, Ohta A, Sako Y, Yamagishi Y, Murakami H, Suga H. Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. ACS Chem Biol. 2008;3 (2):120–9.
126. Palmer E, Liu H, Khan F, Taussig MJ, He M. Enhanced cell-free protein expression by fusion with immunoglobulin Ckappa domain. Protein Sci. 2006;15(12):2842–6.
127. He M, Taussig MJ. Single step generation of protein arrays from DNA by cell-free expression and *in situ* immobilisation (PISA method). Nucleic Acids Res. 2001;29(15):E73–3.
128. Angenendt P, Kreutzberger J, Glokler J, Hoheisel JD. Generation of high density protein microarrays by cell-free *in situ* expression of unpurified PCR products. Mol Cell Proteomics. 2006;5(9):1658–66.
129. Ramachandran N, Hainsworth E, Demirkan G, LaBaer J. On-chip protein synthesis for making microarrays. Methods Mol Biol. 2006;328:1–14.
130. Tao SC, Zhu H. Protein chip fabrication by capture of nascent polypeptides. Nat Biotechnol. 2006;24(10):1253–4.
131. Ray S, Mehta G, Srivastava S. Label-free detection techniques for protein microarrays: prospects, merits and challenges. Proteomics. 2010;10(4):731–48.
132. Schweitzer B, Wiltshire S, Lambert J, O'Malley S, Kukanskis K, Zhu Z, *et al.* Inaugural article: immunoassays with rolling circle DNA amplification: a versatile platform for ultrasensitive antigen detection. Proc Natl Acad Sci USA. 2000;97 (18):10113–9.
133. Schweitzer B, Roberts S, Grimwade B, Shao W, Wang M, Fu Q, *et al.* Multiplexed protein profiling on microarrays by rolling-circle amplification. Nat Biotechnol. 2002;20(4):359–65.
134. Martin K, Steinberg TH, Cooley LA, Gee KR, Beechem JM, Patton WF. Quantitative analysis of protein phosphorylation status and protein kinase activity on microarrays using a novel fluorescent phosphorylation sensor dye. Proteomics. 2003;3 (7):1244–55.
135. Ojida A, Mito-oka Y, Sada K, Hamachi I. Molecular recognition and fluorescence sensing of monophosphorylated peptides in aqueous solution by bis(zinc(II)-dipicolylamine)-based artificial receptors. J Am Chem Soc. 2004;126(8):2454–63.
136. Salisbury CM, Maly DJ, Ellman JA. Peptide microarrays for the determination of protease substrate specificity. J Am Chem Soc. 2002;124(50):14868–70.
137. Zhu Q, Uttamchandani M, Li D, Lesaicherre ML, Yao SQ. Enzymatic profiling system in a small-molecule microarray. Org Lett. 2003;5(8):1257–60.
138. Chen GY, Uttamchandani M, Zhu Q, Wang G, Yao SQ. Developing a strategy for activity-based detection of enzymes in a protein microarray. Chembiochem. 2003;4(4):336–9.
139. Takahashi M, Nokihara K, Mihara H. Construction of a protein-detection system using a loop peptide library with a fluorescence label. Chem Biol. 2003;10(1):53–60.
140. Usui K, Takahashi M, Nokihara K, Mihara H. Peptide arrays with designed alpha-helical structures for characterization of proteins from FRET fingerprint patterns. Mol Divers. 2004;8 (3):209–18.
141. Schweitzer B, Predki P, Snyder M. Microarrays to characterize protein interactions on a whole-proteome scale. Proteomics. 2003;3(11):2190–9.
142. Yu X, Xu D, Cheng Q. Label-free detection methods for protein microarrays. Proteomics. 2006;6(20):5493–503.
143. Chen L, Fatima S, Peng J, Leng X. SELDI protein chip technology for the detection of serum biomarkers for liver disease. Protein Pept Lett. 2009;16(5):467–72.
144. Poon TC. Opportunities and limitations of SELDI-TOF-MS in biomedical research: practical advices. Expert Rev Proteomics. 2007;4(1):51–65.
145. Kozak KR, Su F, Whitelegge JP, Faull K, Reddy S, Farias-Eisner R. Characterization of serum biomarkers for detection of early stage ovarian cancer. Proteomics. 2005;5(17):4589–96.
146. Malik G, Rojahn E, Ward MD, Gretzer MB, Partin AW, Semmes OJ, *et al.* SELDI protein profiling of dunning R-3327 derived cell lines: identification of molecular markers of prostate cancer progression. Prostate. 2007;67(14):1565–75.
147. Lebrecht A, Boehm D, Schmidt M, Koelbl H, Grus FH. Surface-enhanced laser desorption/ionisation time-of-flight mass spectrometry to detect breast cancer markers in tears and serum. Cancer Genomics Proteomics. 2009;6(2):75–83.
148. Schmidt M, Hasenclever D, Schaeffer M, Boehm D, Cotarelo C, Steiner E, *et al.* Prognostic effect of epithelial cell adhesion molecule overexpression in untreated node-negative breast cancer. Clin Cancer Res. 2008;14(18):5849–55.
149. Albrethsen J, Bogebo R, Gammeltoft S, Olsen J, Winther B, Raskov H. Upregulated expression of human neutrophil peptides 1, 2 and 3 (HNP 1-3) in colon cancer serum and tumours: a biomarker study. BMC Cancer. 2005;5:8.
150. Orvisky E, Drake SK, Martin BM, Abdel-Hamid M, Ressom HW, Varghese RS, *et al.* Enrichment of low molecular weight fraction of serum for MS analysis of peptides associated with hepatocellular carcinoma. Proteomics. 2006;6(9):2895–902.

151. McDonnell JM. Surface plasmon resonance: towards an understanding of the mechanisms of biological molecular recognition. Curr Opin Chem Biol. 2001;5(5):572–7.
152. Jung SO, Ro HS, Kho BH, Shin YB, Kim MG, Chung BH. Surface plasmon resonance imaging-based protein arrays for high-throughput screening of protein–protein interaction inhibitors. Proteomics. 2005;5(17):4427–31.
153. Campbell CT, Kim G. SPR microscopy and its applications to high-throughput analyses of biomolecular binding events and their kinetics. Biomaterials. 2007;28(15):2380–92.
154. Watanabe S, Usui K, Tomizaki KY, Kajikawa K, Mihara H. Anomalous reflection of gold applicable for a practical protein-detecting chip platform. Mol Biosyst. 2005;1(5–6):363–5.
155. Watanabe S, Tomizaki KY, Takahashi T, Usui K, Kajikawa K, Mihara H. Interactions between peptides containing nucleobase amino acids and T7 phages displaying S. cerevisiae proteins. Biopolymers. 2007;88(2):131–40.
156. Kaushansky A, Gordus A, Budnik BA, Lane WS, Rush J, MacBeath G. System-wide investigation of ErbB4 reveals 19 sites of Tyr phosphorylation that are unusually selective in their recruitment properties. Chem Biol. 2008;15(8):808–17.
157. Gong W, He K, Covington M, Dinesh-Kumar SP, Snyder M, Harmer SL, *et al.* The development of protein microarrays and their applications in DNA-protein and protein–protein interaction analyses of Arabidopsis transcription factors. Mol Plant. 2008;1(1):27–41.
158. Zhu H, Hu S, Jona G, Zhu X, Kreiswirth N, Willey BM, *et al.* Severe acute respiratory syndrome diagnostics using a coronavirus protein microarray. Proc Natl Acad Sci USA. 2006;103(11):4011–6.
159. Huang J, Zhu H, Haggarty SJ, Spring DR, Hwang H, Jin F, *et al.* Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips. Proc Natl Acad Sci USA. 2004;101(47):16594–9.
160. Sokolov BP, Cadet JL. Methamphetamine causes alterations in the MAP kinase-related pathways in the brains of mice that display increased aggressiveness. Neuropsychopharmacology. 2006;31(5):956–66.
161. Yam CM, Deluge M, Tang D, Kumar A, Cai C. Preparation, characterization, resistance to protein adsorption, and specific avidin-biotin binding of poly(amidoamine) dendrimers functionalized with oligo(ethylene glycol) on gold. J Colloid Interface Sci. 2006;296(1):118–30.
162. Cha T, Guo A, Zhu XY. Enzymatic activity on a chip: the critical role of protein orientation. Proteomics. 2005;5(2):416–9.
163. George N, Pick H, Vogel H, Johnsson N, Johnsson K. Specific labeling of cell surface proteins with chemically diverse compounds. J Am Chem Soc. 2004;126(29):8896–7.
164. Juillerat A, Heinis C, Sielaff I, Barnikow J, Jaccard H, Kunz B, *et al.* Engineering substrate specificity of O6-alkylguanine-DNA alkyltransferase for specific protein labeling in living cells. Chembiochem. 2005;6(7):1263–9.
165. Nath N, Hurst R, Hook B, Meisenheimer P, Zhao KQ, Nassif N, *et al.* Improving protein array performance: focus on washing and storage conditions. J Proteome Res. 2008;7(10):4475–82.

Biological Activities of Natural and Engineered Cyclotides, a Novel Molecular Scaffold for Peptide-Based Therapeutics

Angie E. Garcia and Julio A. Camarero*

Department of Pharmacology and Pharmaceutical Sciences, School of Pharmacy, University of Southern California, Los Angeles, CA 90033, USA

Abstract: Cyclotides are a growing family of large plant-derived backbone-cyclized polypeptides (≈30 amino acids long) that share a disulfide-stabilized core characterized by an unusual knotted structure. Their unique circular backbone topology and knotted arrangement of three disulfide bonds makes them exceptionally stable to thermal, chemical, and enzymatic degradation compared to other peptides of similar size. Currently more than 100 sequences of different cyclotides have been characterized and the number is expected to increase dramatically in the coming years. Considering their stability, biological activities and ability to cross the cell membrane, cyclotides can be exploited to develop new peptide-based drugs with high potential for success. The cyclotide scaffold can be engineered or evolved using molecular evolution to inhibit protein-protein interactions implicated in cancer and other human diseases, or design new antimicrobials. The present review reports the biological diversity and therapeutic potential of natural and engineered cyclotides.

Keywords: Cyclic peptides, Cyclotides, Intein Kalata B1, MCoTI-I/II, Protein splicing, Protein engineering.

INTRODUCTION

Head-to-tail or backbone-cyclized peptides are present throughout nature from bacteria to animals. Bacteria and fungi express numerous backbone-cyclized peptides that are currently in use as therapeutic agents [1]. For example, cyclosporin A is a fungal peptide with potent immunosuppressive properties and is used to treat organ transplant patients [2]. Daptomycin, is a 13-amino acid cyclic lipopeptide with a decanoyl side chain isolated from *Streptomyces roseosporus* that has recently been approved to treat infections against Gram-positive organisms, including multi-resistant strains [3]. In animals, the only known circular peptides are θ-defensins, which are expressed in blood leukocytes and bone marrow of Old World monkeys [4, 5]. θ-Defensins are antimicrobial peptides with broad-spectrum activities against bacteria, fungi, and viruses [6-8]. Backbone cyclized peptides have been also found in plants [9]. Sunflower trypsin inhibitor 1 (SFTI-1) for example is a bicyclic 14-residue long peptide found in sunflower seeds. SFTI-1 is the most potent known naturally occurring Bowman-Birk trypsin inhibitor [10]. Cyclotides, a novel family of small globular backbone-cyclized micro-proteins (≈ 30-residues long), are also naturally found in plants. Here, we review the properties of cyclotides and the latest developments in the use of the cyclotide scaffold to design novel peptide-based therapeutics.

CYCLOTIDES, A NOVEL ULTRASTABLE MOLECULAR SCAFFOLD

Cyclotides are small globular micro-proteins with a unique head-to-tail cyclized backbone, which is stabilized by three disulfide bonds (Fig. **1**). Currently, over 140 sequences have been identified in the plant species *Rubiaceae*, *Violaceae*, and *Cucurbitaceae* [11]. Natural cyclotides have various activities including insecticidal [12, 13], uterotonic [14], anti-HIV [15], antimicrobial [16, 17], antitumor [18], antihelminthic [19, 20] and have been reported to cross cell membranes [21]. Their insecticidal and antihelminthic properties suggest that they may function as defense molecules in plants.

Cyclotides share a unique head-to-tail circular knotted topology of three disulfide bridges, with one disulfide penetrating through a macrocycle formed by the two other disulfides and inter-connecting peptide backbones, forming what is called a cyclic cystine knot (CCK) motif (Fig. **1**). The CCK topology is responsible for the high stability of cyclotides to enzymatic, thermal and chemical degradation [22]. The cyclotide family is divided into three structurally distinct families, Möbius, bracelet, and trypsin inhibitor subfamilies (Fig. **1**). Möbius cyclotides are distinguished from bracelet cyclotides by the presence of a *cis*-Pro residue in loop 5. Trypsin inhibitor cyclotides have very different primary structures from Möbius and bracelet cyclotides, but retain the conserved cystine knot motif. Trypsin inhibitor cyclotides share a high sequence homology with related cystine-knot trypsin inhibitors found in squash such as EETI-II (*Ecballium elaterium* trypsin inhibitor II), and in fact can be considered cyclized homologs of these protease inhibitors. Thus, cyclotides can be considered natural combinatorial peptide libraries structurally constrained by the cystine-knot scaffold [23] and head-to-tail cyclization but are permissive of hypermutation of essentially all residues with the exception of the strictly conserved cysteines that comprise the knot [24-27].

Hence, cyclotides form a unique family of structurally-related peptides that possess remarkable stability due to the cystine knot, a small size making them readily accessible to chemical synthesis, and an excellent tolerance to sequence variations. Moreover, the first cyclotide to be discovered, kalata B1, is an orally effective uterotonic [14]. Intriguingly, the cyclotide MCoTI-II has also been shown to cross cell

*Address correspondence to this author at the Department of Pharmacology and Pharmaceutical Sciences, University of Southern California, 1985 Zonal Avenue, PSC 616, Los Angeles, CA 90033, USA; Tel: 323-442-1417; Fax: 323-224-7473; E-mail: jcamarer@usc.edu

1874-4672/10 $55.00+.00

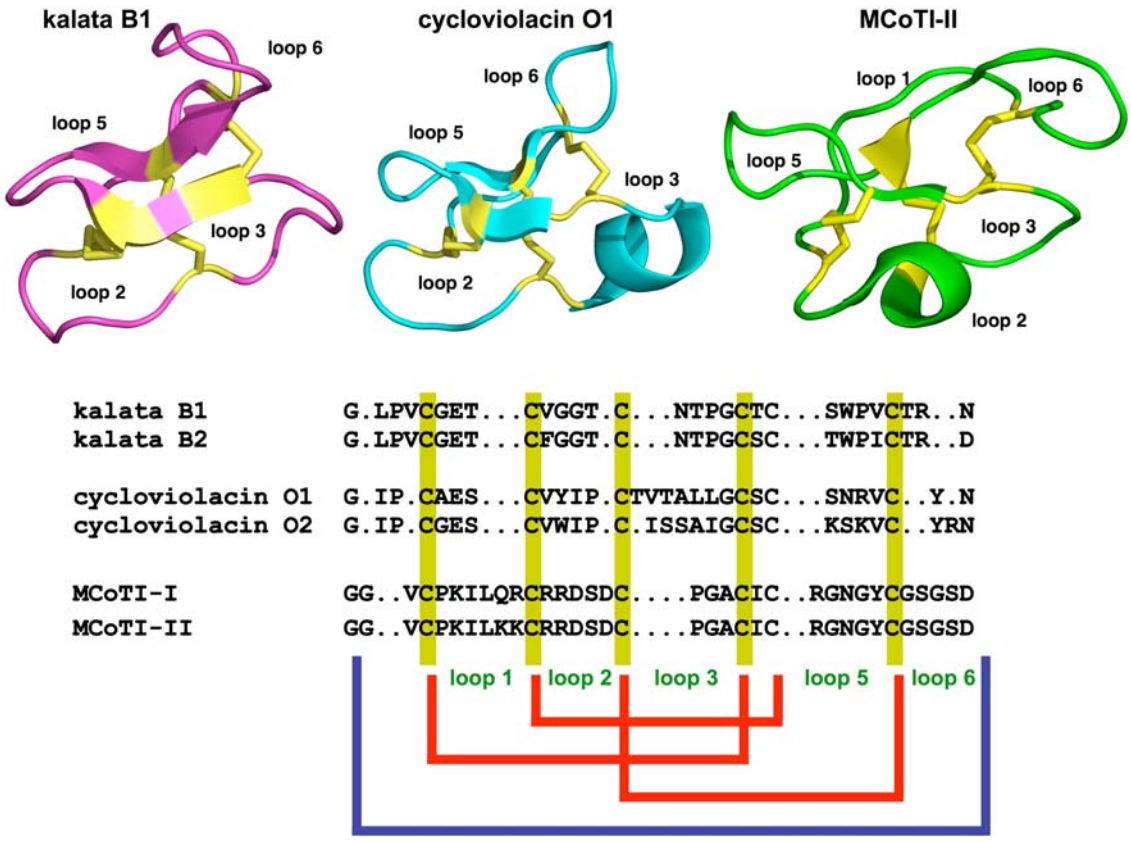


Fig. (1). Primary and tertiary structures of representative cyclotides from the bracelet (kalata B1; pdb ID: 1NB1 [108]), Möbius (cycloviolacin O1, pdb ID: 1NBJ [108]), and trypsin inhibitor (MCoTI-II, pdb ID: 1IB9 [109]) subfamilies. Conserved cysteine residues and disulfide bonds are shown in yellow. The blue line denotes the circular backbone.

membranes through macropinocytosis [21]. We have also recently found that MCoTI-I has similar cellular-uptake properties to MCoTI-II (unpublished results). All of these features make cyclotides ideal tools for the development of a total novel class of peptide-based therapeutics.

CYCLOTIDE BIOSYNTHESIS

Cyclotides are ribosomally synthesized as precursor proteins, which consist of an endoplasmic reticulum (ER)-targeting sequence, a pro-region, a highly conserved N-terminal repeat (NTR) region, a mature cyclotide domain, and a C-terminal tail (Fig. **2**). The combined NTR-cyclotide segment may contain one copy of the cyclotide sequence or there may be multiple copies of the same or different cyclotide sequences separated by additional NTR sequences. The precursor undergoes post-translational processing to generate a circular peptide by a mechanism that has not been completely elucidated yet [28, 29]. It has been hypothesized that a conserved Asn (or Asp) at the C-terminal cleavage site may be a recognition site by asparaginyl endoproteinase (AEP) for cyclization of the peptide in plants [28, 29]. AEP has been shown to be involved in the post-translational processing of concanavalin A from the jackbean [30], and therefore, it is possible this enzyme may be involved in the cyclization of other plant peptides. Studies using transgenic plants that express a cyclotide precursor have demonstrated the involvement of AEP and requirement for the asparagine residue in the cyclotide sequence [28, 29]. The authors showed that inhibition of AEP led to a decrease in the amount of cyclic product and an accumulation of linear peptides that were transiently expressed [29]. In addition, a complementary study showed that mutation of the asparagine residue or truncation of the conserved C-terminal tripeptide in transgenic plants resulted in no circular peptide production [28].

BIOLOGICAL ACTIVITIES OF NATURALLY OCCURRING CYCLOTIDES

The natural function of cyclotides appears to be in protection of plants against insects [12, 13], nematodes [20, 25], and mollusks [31]. Studies have demonstrated that cyclotides can suppress the growth and development of insect and nematode larvae. Various other studies have also shown cyclotides have antimicrobial, hemolytic, uterotonic, and anti-HIV activities. Much of these activities likely involve interaction of the cyclotide with membranes, although the mechanism of action is not totally well understood.

The first cyclotide discovered, kalata B1, was identified in the plant *Oldenlandia affinis* in central Africa in the 1960's [32]. This plant was used by the natives to make a tea extract that was used to accelerate childbirth during labor [14, 33]. The main active ingredient in the tea extract was found to be a peptide that was named kalata B1, after the local name for the native medicine. The uterotonic properties

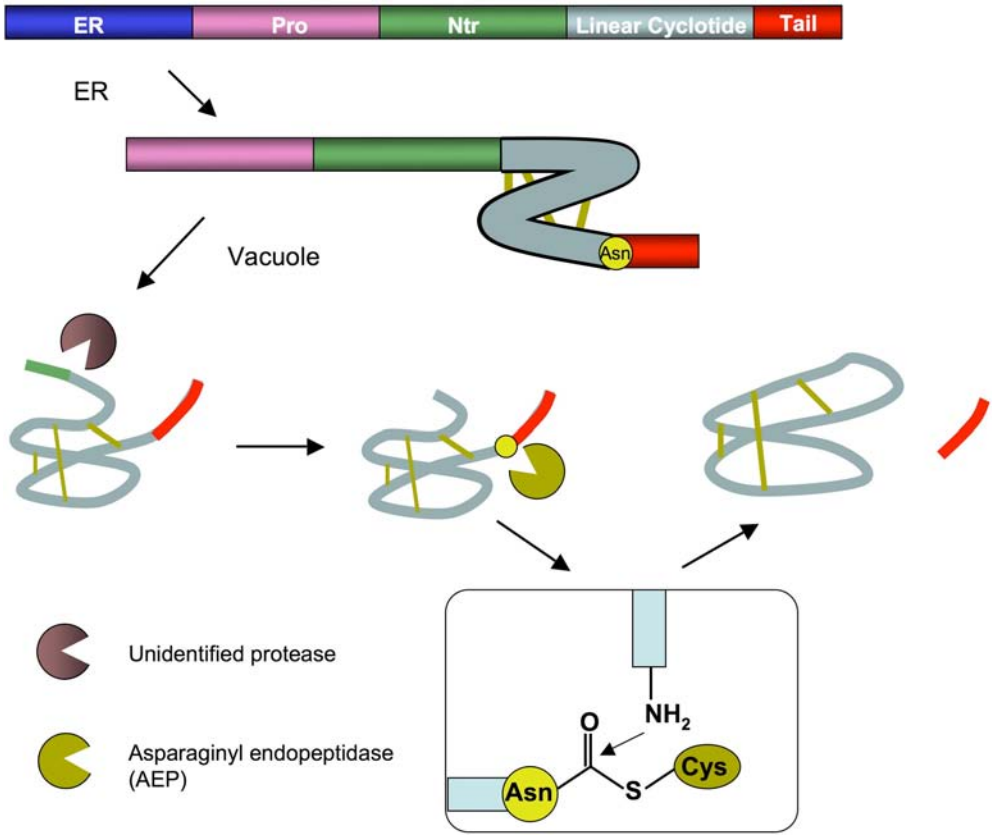


Fig. (2). Schematic representation of the putative mechanism of protease-catalyzed cyclization for cyclotides [28, 29]. The prototypic linear precursor protein (top) comprises an endoplasmic reticulum (ER)-targeting sequence (dark blue), a pro-region (purple), an N-terminal repeat region (Ntr; green), a cyclotide domain (light grey) and a C-terminal tail (red). It also features a conserved asparagine (yellow) at the C-terminal cleavage point of the cyclotide domain. The precursor is processed in the ER and vacuole, disulfide bonds are formed (yellow), and a range of unidentified proteases (brown) trim the precursor. In the final stage, the active-site cysteine of an AEP (yellow) displaces the C-terminal tail to form an enzyme-acyl intermediate (boxed). This intermediate is then attacked by the cyclotide N-terminal glycine to form the mature cyclic peptide. Figure taken from reference [68].

of kalata B1 indicated that the peptide was orally bioavailable. The complete sequence, Cys-knotted arrangement, and cyclic nature of kalata B1 was determined 25 years after its original discovery [34]. Since the discovery of kalata B1, many more related cyclotides have been discovered and found to have various biological activities [35, 36].

Cyclotides were initially hypothesized to have antimicrobial activities based on the presence of hydrophilic and hydrophobic patches, which give an amphipathic character similar to classical antimicrobial peptides. The antimicrobial activities of cyclotides have been reported by two groups with conflicting results on the potency of kalata B1 against *Escherichia coli* and *Staphylococcus aureus.* In one study, kalata B1 was active against *S. aureus*, but not *E. coli* [16], and in the second study, the peptide had the reverse effect [17]. This is likely due to the technical differences in the experiments. Although kalata cyclotides are amphipathic, the overall charge is close to zero at neutral pH, making it unlikely that they interact with bacterial membranes electrostatically similar to classical cationic antimicrobial peptides [37]. Further studies are necessary to investigate the mechanisms of antimicrobial action given the growing occurrence of antibiotic resistance by microorganisms.

The anti-HIV properties of cyclotides have been extensively studied [15, 38-40]. They appear to mainly act by inhibiting viral entry into host cells as studies have shown a dose-dependent increase in cytoprotection [15]. This suggests the peptides may block binding or fusion of the virus, but the mechanism remains unclear. In general, cyclotide bioactivities appear to involve interactions with membranes and, therefore, this may be a mechanism for anti-HIV activities, by preventing fusion of the viral and host cell membranes. Studies have shown cyclotides can bind to model lipid membranes by surface plasmon resonance [41], and that binding occurs mainly through the peptide hydrophobic patches exposed on the surface [42-44]. This suggests membrane binding may be one mode for cyclotide activity against microorganisms.

Cyclotide interactions with membranes have also been suggested as their mechanism for cytotoxic activity. Studies have demonstrated antitumor activities of cyclotides, which were selective against cancer cell lines and solid tumors compared to normal mammalian cells [18, 45, 46]. Cancer cells differ from normal cells in the lipid and glycoprotein composition, which alters the overall net charge. The different potencies between cyclotide cytotoxicity are related to

the three-dimensional structure as well as specific amino acid residues within the sequence [46, 47].

In addition to having antimicrobial and antitumor activities, some cyclotides have been found to cause extensive hemolysis of human and rat erythrocytes [16, 40, 48]. The cyclotide kalata B1 has strong hemolytic activity, although this can be eliminated by mutation to Ala of any one of eight residues located in the bioactive face of the molecule [25]. A more recent study has also shown that the hemolytic activity of kalata B1 could be reduced or completely eliminated by mutation to Lys of any of the residues involved in either the bioactive or hydrophobic faces of kalata B1 [49]. The same study showed that the hemolytic, insecticidal and nematocidal activities of the different kalata B1 mutants were correlated, indicating there may be a common mechanism involving a cyclotide-membrane interaction [49]. The same authors have also shown recently the all D-analogue of kalata had similar nematicidal activity than the native peptide, thus corroborating that the biological activity is not mediated by a cellular receptor [20]. In contrast, other cyclotides such as MCoTI-cyclotides have shown little or no hemolytic activity, thus demonstrating the diversity of cyclotide properties.

Another cyclotide with interesting biological activity is cyclopsychotride (Cpt) A. Cpt A is a natural cyclotide obtained from the organic extract of the tropical plant *Psychotropia longipes* that has been reported to have neurotensin inhibition properties [50]. Cpt A was able to inhibit neurotensin binding to its receptor to HT-29 cell membranes with an $IC_{50} \approx 3$ μM and increase intracellular Ca^{2+} levels in a concentration-dependent manner, which could not be blocked by neurotensin antagonists [50]. Cpt A, however, showed a similar activity in two unrelated cell lines that did not express neurotensin receptors indicating that the mechanism of action is unlikely to be mediated through an interaction with the neurotensin receptor [50].

CHEMICAL SYNTHESIS OF CYCLOTIDES

Cyclotides are small peptides, approximately 30 amino acids long, and therefore can be readily synthesized by chemical methods using solid-phase peptide synthesis [51]. Chemical synthesis using a solid-phase approach has been utilized to generate native cyclotide structures as well as grafted analogues [52-56]. This method uses an intramolecular native chemical ligation [57], in which the peptide sequence contains an N-terminal cysteine and an α-thioester group at the C-terminus [58-60]. Both tert-butyloxyxarbonyl (Boc)- and 9-fluorenyloxycarbonyl (Fmoc)-based chemistries have been used to incorporate C-terminal thioesters during chain assembly (Boc) [61-63] or using a safety-catch based linkers (Fmoc) [60, 64-67]. Once the peptide is cleaved from the resin, both cyclization and folding are carried out in a single pot reaction.

RECOMBINANT EXPRESSION OF CYCLOTIDES

Cyclotides have also been produced recombinantly in bacteria through intramolecular native chemical ligation (see above) by using a modified protein splicing unit or intein (Fig. **3**) (see reference [68] for a recent review). This method can generate folded cyclotides either *in vivo* or *in vitro* using standard bacterial expression systems [26, 69, 70]. Inteins are internal self-processing domains that undergo post-translational processing to splice together flanking external domains (exteins) [71]. The approach uses a modified intein fused to the C-terminus of the cyclotide sequence to allow the formation of an α-thioester at the C-terminus of recombinant polypeptides. To obtain the required N-terminal cysteine for cyclization, the peptide can be expressed with an N-terminal leading peptide signal, which can be cleaved either *in vivo* or *in vitro* by proteolysis or auto-proteolysis [68]. The simplest way to accomplish this is to introduce a Cys downstream of the initiating Met residue. Once the translation step is completed, the endogeneous methionyl aminopeptidases (MAP) removes the Met residue, thereby generating *in vivo* an N-terminal Cys residue [72-76]. The N-terminal Cys can then capture the reactive thioester in an intramolecular fashion to form a backbone-cyclized polypeptide (Fig. **3**). Additional methods to generate an N-terminal cysteine have used exogenous proteases to cleave the leading signal after purification or *in vivo* by co-expressing the protease [77]. For example, the protease Factor Xa has been used to remove an N-terminal recognition sequence prior to a cysteine residue [59, 78]. Other proteases that have been used for this task include ubiquitin C-terminal hydrolase [79, 80], tobacco etch virus (TEV) protease [77], enterokinase [81] and thrombin [82]. The N-terminal pelB leader sequence has been used recently to direct newly synthesized fusion proteins to the *E. coli* periplasmic space where the corresponding endogenous leader peptidases [83, 84] can generate the desired N-terminal cysteine-containing protein fragment [85]. Besides proteases, protein splicing has also been used to produce recombinant N-terminal Cys-containing polypeptides. Some inteins can be modified in such a way that cleavage at the C-terminal splice junction can be accomplished in a pH- and temperature-dependent fashion [86-88].

Intein-mediated backbone cyclization of polypeptides has also been recently used for the biosynthesis of the Bowman-Birk inhibitor SFTI-1 [89]. The biosynthesis of other cyclic peptides such as backbone-cyclized α-defensins and naturally occurring θ-defensins is currently underway in our laboratory.

Another approach to generate cyclic peptides *in vivo* is by protein trans-splicing. This approach utilizes a self-processing intein that is split into two fragments, an N-intein and a C-intein. This method has not been applied yet for the biosynthesis of cyclotides, but has been used to produce other natural cyclic peptides and genetically-encoded libraries of small cyclic peptides [90, 91]. It should be noted, however, that these systems require the presence of specific amino acid residues at both intein-extein junctions for efficient protein splicing to occur [90, 92, 93].

DESIGNING CYCLOTIDES WITH NOVEL BIOLOGICAL ACTIVITIES

The unique properties associated with the cyclotide scaffold make them extremely valuable tools in drug discovery. There are several studies that have used the cyclotide

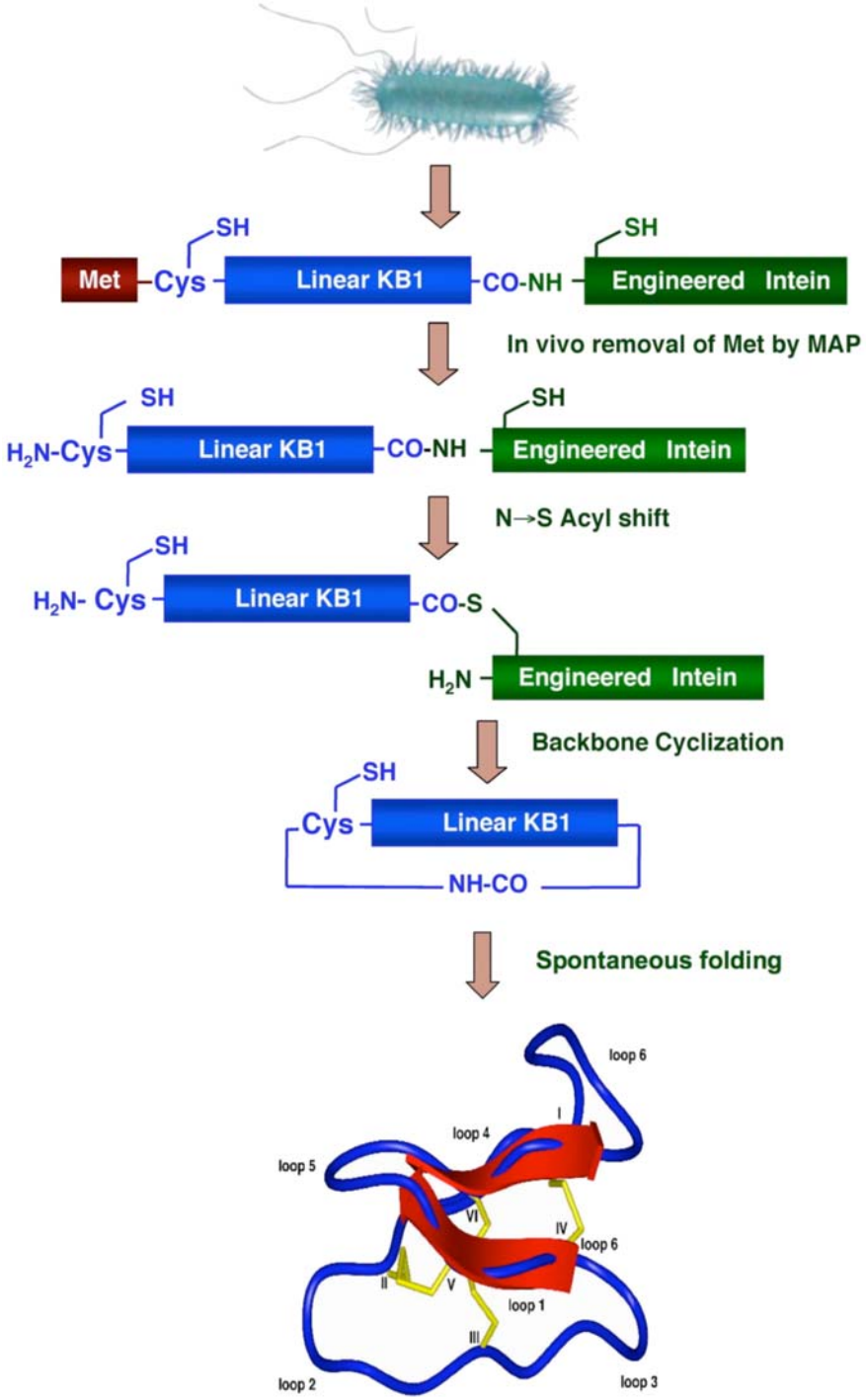


Fig. (3). Intein-mediated backbone cyclization for the biosynthesis of cyclotides kalata B1 (kB1) and MCoTI-II in *E. coli* cells [69, 70]. The backbone cyclization of the linear cyclotide precursor is mediated by a modified protein splicing unit or intein. The cyclized product then folds spontaneously in the bacterial cytoplasm.

molecular scaffold to graft peptide sequences and to generate libraries for the purpose of engineering cyclotides with novel biological functions (Fig. **4**).

The plasticity of the cyclotide framework was first demonstrated by substituting hydrophobic residues in loop 5 of kalata B1 with polar and charged residues [24]. The mutated cyclotides retained the native fold of kalata B1, but were no longer hemolytic [24]. This showed that cyclotides were amendable to sequence changes, and interestingly, can be modified to change their biological functions.

The potential of grafted cyclotides was first demonstrated in a study aimed to develop novel anticancer peptide-based therapeutics [94]. In this work a peptide antagonist of angiogenesis was grafted into various loops of the kalata B1 scaffold [94]. The grafted cyclotide containing the vascular endothelial growth factor A (VEGF-A) antagonist sequence in

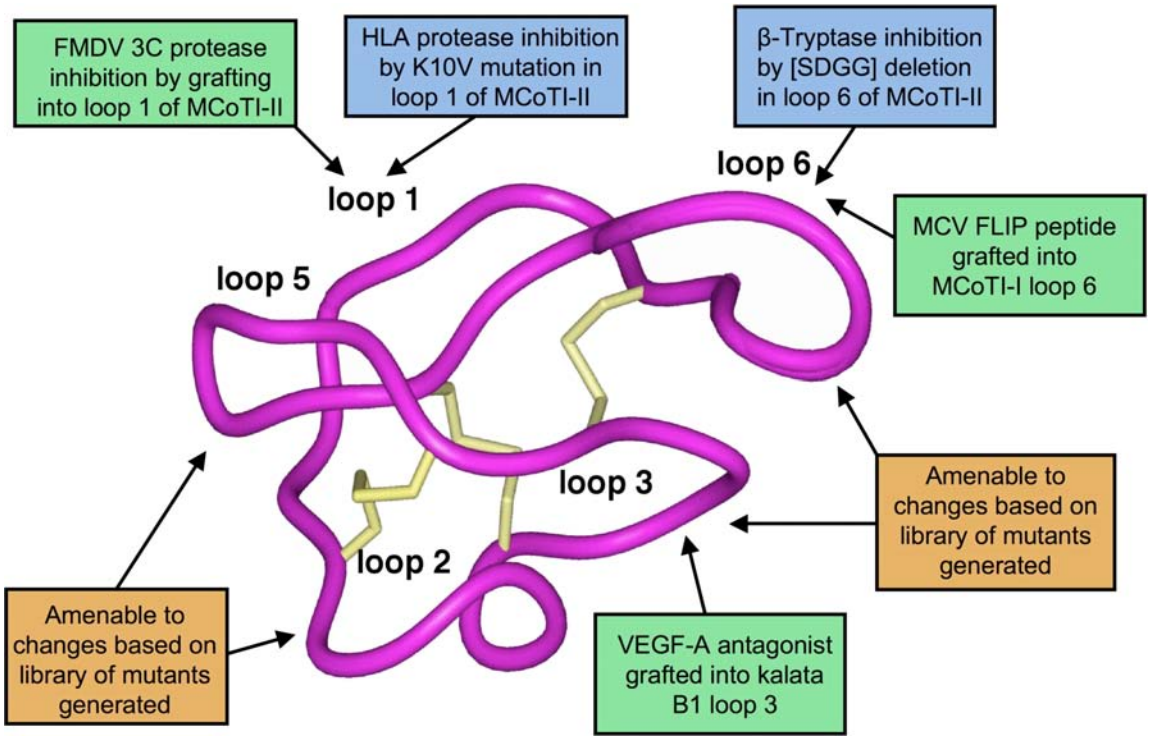


Fig. (4). Using the cyclotide molecular scaffold for drug design. Summary of the changes engineered into the cyclotide framework to introduce new biological functions. Peptide sequences have been successfully grafted into loop 3 of kalata B1 [94] and loops 1 and 6 of MCoTI-II [95] and MCoTI-I (unpublished results), respectively. Cyclotide based libraries have been generated using loops 2, 3, 5 and 6. The cyclotide structure used in the figure corresponds to MCoTI-II (pdb ID: 1IB9 [109]).

loop 3 was found to adopt a CCK native fold and was biologically active at low micromolar concentrations. Additionally, the grafted cyclotide showed increased resistance to degradation in human serum. This study demonstrates the possibility of using the cyclotide scaffold to stabilize bioactive peptide epitopes, which may normally get degraded.

The utility of the cyclotide scaffold in drug design has also been recently shown by engineering non-native activities into the cyclotide MCoTI-II. MCoTI-II is a naturally occurring trypsin inhibitor ($K_i \approx 20$ pM, which is the dissociation constant for inhibitor binding) found in the seeds of *Momordica cochinchinensis*, a tropical plant from the squash family. Mutation of the P1 residue in the active loop (loop 1, see Fig. **1**) of the cyclotide produced several MCoTI-II analogs with different specificities towards different proteases [95]. Interestingly, several analogs showed activity against the foot-and-mouth-disease virus (FMDV) 3C protease, a Cys protease key for viral replication, in the low micromolar range [95]. This is the first reported peptide-based inhibitor for this protease and although the potency was relatively low, this study demonstrates the potential of using MCoTI-based cyclotides for designing novel protease inhibitors [95].

In a more recent study, the same authors also generated inhibitors of the serine proteases β-tryptase and human leukocyte elastase (HLE) using the backbone of MCoTI-II [96]. β-Tryptase is implicated in allergic and inflammatory disorders, and HLE has been associated with respiratory and pulmonary disorders. Replacing the P1 residue in loop 1 produced several MCoTI-II mutants (K6A and K6V) with activity against HLE with K_i values of 20-30 nM [96] and K_i values against trypsin above 1 μM. Removal of the SDGG peptide segment in loop 6 yielded a ß-tryptase inhibitor with a $K_i \approx 10$ nM without significantly altering the three-dimensional structure as determined by NMR [96]. The authors hypothesized that deletion of the aspartic acid residue in MCoTI-II should improve activity by removal of repulsive electrostatic interactions with β-tryptase thus improving the inhibitory constant against β-tryptase 160-fold when compared to the wild-type MCoTI-II.

In addition to displaying biological activities, MCoTI-based peptides have also been shown to cross cell membranes in macrophage and breast cancer cell lines through macropinocytosis [21]. We have also found that grafting of a helix region from the molluscum contagiosum virus (MCV) FLICE-inhibitory protein (FLIP) into loop 6 of MCoTI-I yielded a folded cyclotide able to cross cell membranes and trigger apoptosis of virally infected cells (unpublished results), thus indicating that loop 6 may be used for grafting purposes without affecting cellular-uptake.

MCoTI cyclotides share a high sequence homology with related cystine-knot trypsin inhibitors found in squash, and could be considered cyclized homologs of these protease inhibitors. Squash cystine-knot trypsin inhibitors have also successfully been used to graft biological activities. For example, the RGD sequence was grafted into loop 1 of EETI-II yielding an EETI-II analog with platelet inhibitory activity [97]. The engineered proteins were much more potent in inhibiting platelet aggregation than the linear grafted pep-

tides, thus highlighting the importance of grafting an epitope into a stable peptide-scaffold. These highly stable peptides could have clinical use for the treatment of patients with acute coronary syndrome, for example.

Additionally, the cyclotide scaffold can be engineered or evolved through molecular evolution techniques to selectively bind extracellular receptors such as G protein-coupled receptors to promote or block cell signaling [98]. These data demonstrate the versatility of the cyclotide scaffold and highlight the extraordinary pharmacological properties of MCoTI-cyclotides and related linear knottins, thus confirming the potential of Cys-knotted polypeptide scaffolds in peptide-based drug discovery [27].

SCREENING OF CYCLOTIDE-BASED LIBRARIES

The ability to create cyclic polypeptides *in vivo* opens up the intriguing possibility of generating large libraries of cyclic polypeptides. Thus, libraries of genetically-encoded cyclic polypeptides containing billions of members can be readily generated using standard molecular recombinant tools. This tremendous molecular diversity allows one to perform selection strategies mimicking the evolutionary processes found in nature.

There are several examples where *in vivo* generated libraries of cyclic peptides have been used for rapid selection of biologically active peptides. For example protein trans-splicing has been used by several groups for the generation of libraries of small cyclic peptides ($\leq$ 8 residues) in bacterial and mammalian cells [90, 92, 93, 99-101]. It should be pointed out, however, that the use of protein trans-splicing requires particular amino acids at the intein-extein junctions for efficient trans-splicing [93]. These requirements usually depend on the type of split-intein used, and seriously limit the diversity of the libraries that can be generated when using small cyclic peptide templates. Our group has recently demonstrated the expression in *E. coli* of libraries based on the cyclic peptide SFTI-1 (a backbone cyclized Bowman-Birk trypsin inhibitor) using an intein-mediated backbone cyclization approach (see above), which allows the biosynthesis of backbone-cyclized polypeptides without any sequence requirement limitation [89].

The use of small cyclic peptides, however, which technically can be considered a single closed loop (or 2 loops in the case of SFTI-1), could also limit the potency of the peptides selected, especially when targeting protein-protein interactions involving large binding surfaces. In these cases the use of peptide templates such as cyclotides with multiple variable loops could facilitate the selection of peptides with higher affinities.

The potential for generating cyclotide libraries was first explored by our group using the kalata B1 scaffold [69]. In this work wild-type and several mutants of kalata B1 were biosynthesized using an intramolecular native chemical ligation facilitated by a modified protein splicing unit. In this work, six different linear versions of kalata B1 were generated and expressed in *E. coli* as fusions to a modified version of the yeast vacuolar membrane ATPase (VMA) intein. Results demonstrated *in vitro* folding and cyclization of kalata B1 to varying degrees depending on which of the six native cysteine residues was at the N-terminus after cleavage of the initiation methionine by endogenous MAP. Cleavage and efficient cyclization of the different linear precursors did not occur equally, suggesting the amino acid residues near the intein as well as the predisposition to adopt a native fold of the corresponding linear precursor may determine the efficiency of the cleavage/cyclization step [69]. This information was used to express a small library based on the kalata B1 scaffold. This library was cyclized *in vitro* by incubation with a redox buffer containing reduced glutathione (GSH) as a thiol co-factor, thus mimicking the intracellular conditions, where GSH is the most abundant thiol co-factor. The use of GSH allows the cyclization and folding to happen in one step [69, 89]. Analysis of the cyclization/folding reaction by HPLC and mass spectrometry revealed that all the members of the kalata B1 based library were expressed and processed with similar yields to give the corresponding natively folded cyclotides [69].

More recently, we have also reported the biosynthesis of a genetically encoded library of MCoTI-I based cyclotides in *E. coli* cells [26]. The cyclization/folding of the library was performed either *in vitro,* by incubation with a redox buffer containing glutathione, or by *in vivo* self-processing of the corresponding precursor proteins. The bacterial gyrase A intein from *Mycobacterium xenopus* was used in this work [26]. This intein typically express at higher yields than the yeast VMA intein in *E. coli* expression systems [26]. The peptide libraries were purified and screened for activity using trypsin-immobilized sepharose beads, and then analyzed by HPLC and mass spectrometry. Out of 27 mutations studied, only two mutations, G27P and I22G, negatively affected the folding of the resulting cyclotides. All of the remaining cyclotides were able to fold with similar yields. The K6A mutant, as expected, was not able to bind trypsin. This residue is key for binding to the specificity pocket of trypsin, and can only be replaced by positively charged residues [102]. This mutant was found by NMR to adopt a native cyclotide structure, confirming that the lack of biological activity was due to the mutation and not to the ability to adopt a native fold. It is interesting to note that by modifying the nature of this residue, the specificity of the corresponding MCoTI-cyclotide can be changed to target other proteases [95, 96]. The rest of the MCoTI-based library members were able to bind trypsin, suggesting they were able to adopt a native cyclotide fold and retained biological activity. The affinity of each peptide was analyzed using a competitive trypsin-binding assay. The mutants had a wide range of affinity, some being greater than wild type MCoTI-I. The peptides with less affinity were mostly found in loop 1 and the C-terminal region of loop 6, both well conserved among other squash trypsin inhibitors. Overall, these data describe the structural requirements for correct formation of MCoTI-I and the residues that are key to modulating trypsin binding. To our knowledge, this is the first time that the biosynthesis of a genetically-encoded library of MCoTI-based cyclotides containing a complete suite of amino acid mutants is reported.

The chemical synthesis of a complete suite of Ala mutants for kalata B1 has also been recently reported [25]. In this work all the mutants were fully characterized structurally and functionally. The results indicated that only two of

the mutations explored (W23A and P24A, both located in loop 5, see Fig. **1**) prevented folding [25]. The mutagenesis results obtained in our work with the cyclotide MCoTI-I show similar results highlighting the extreme robustness of the cyclotide scaffold to mutations. These studies show that cyclotides may provide an ideal scaffold for the biosynthesis of large combinatorial libraries inside living bacterial cells. These genetically-encoded libraries can then be screened in-cell for biological activity using high-throughput flow cytometry techniques for the rapid selection of novel biologically active cyclotides [69, 103, 104].

SUMMARY AND CONCLUDING REMARKS

In summary, cyclotides are a novel family of structurally related globular microproteins with a unique head-to-tail cyclized backbone, which is stabilized by three disulfide bonds [27, 105]. The number and positions of cysteine residues are conserved throughout the family, forming what is called cyclic cystine-knot (CCK) motif [35] that acts as a highly stable and versatile scaffold on which 5 hypervariable loops are arranged (Fig. **1**). This CCK framework gives the cyclotides exceptional resistance to thermal and chemical denaturation, and enzymatic degradation. This is particularly important for the development of peptide-based therapeutics with oral bioavailability. In fact, the use of cyclotide-containing plants in indigenous medicine first highlighted the fact that the peptides are resistant to boiling and are apparently orally bioavailable [14, 32, 33]. Some cyclotides have also been shown to cross the cell membrane [21] thus allowing to target intracellular protein interactions, such as that mediated by viral FLIPs to prevent cell-death in virally infected cells (unpublished results). Cyclotides are also medium-sized polypeptides and therefore can be readily synthesized by standard solid-phase peptide-synthesis using either Boc- [54] or Fmoc-based [55] methodologies thus allowing the introduction of non-natural amino acids or other chemical modifications for lead optimization. They can also be encoded within standard cloning vectors and readily expressed in bacteria or animal cells [26, 70], thus making them ideal substrates for molecular evolution strategies to enable generation and selection of compounds with optimal binding and inhibitory characteristics using high throughput cell-based assays [106]. All of these characteristics make cyclotides appear as very promising leads or frameworks for development of peptide-based therapeutics and diagnostics [27, 68, 105, 107]

ACKNOWLEDGEMENTS

This work was supported by funding from the School of Pharmacy at the University of Southern California and by National Institute of Health award GM090323-01 to J.A.C.

ABBREVIATIONS

AEP = asparaginyl endoproteinase

Boc = tert-butyloxyxarbonyl

EETI-II = *Ecballium elaterium* trypsin inhibitor II

FMDV = foot-and-mouth-disease virus

FLIP = FLICE-inhibitory protein

Fmoc = 9-fluorenyloxycarbonyl

GSH = reduced glutathione

HIV = human immunodeficiency virus

HLE = human leukocyte elastase

HPLC = high performance liquid chromatography

MAP = methionyl aminopeptidase

MCoTI = *Momordica cochinchinensis* trypsin inhibitor

MCV = molluscum contagiosum virus

NMR = nuclear magnetic resonance

NTR = N-terminal repeat

SFTI = sunflower trypsin inhibitor

SPPS = solid-phase peptide synthesis

VMA = vacuolar membrane ATPase.

REFERENCES

[1] Mogi, T.; Kita, K. Gramicidin S and polymyxins: the revival of cationic cyclic peptide antibiotics. *Cell. Mol. Life. Sci.*, **2009**, *66(23)*, 3821-3826.

[2] Starzl, T.E.; Klintmalm, G.B.; Porter, K.A.; Iwatsuki, S.; Schroter, G.P. Liver transplantation with use of cyclosporin a and prednisone. *N. Engl. J. Med.*, **1981**, *305(5)*, 266-269.

[3] Enoch, D.A.; Bygott, J.M.; Daly, M.L.; Karas, J.A. Daptomycin. *J. Infect.*, **2007**, *55(3)*, 205-213.

[4] Tang, Y.Q.; Yuan, J.; Osapay, G.; Osapay, K.; Tran, D.; Miller, C.J.; Ouellette, A.J.; Selsted, M.E. A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated alpha-defensins. *Science*, **1999**, *286*, 498-502.

[5] Garcia, A.E.; Osapay, G.; Tran, P.A.; Yuan, J.; Selsted, M.E. Isolation, synthesis, and antimicrobial activities of naturally occurring theta-defensin isoforms from baboon leukocytes. *Infect. Immun.*, **2008**, *76(12)*, 5883-5891.

[6] Gallo, S.A.; Wang, W.; Rawat, S.S.; Jung, G.; Waring, A.J.; Cole, A.M.; Lu, H.; Yan, X.; Daly, N.L.; Craik, D.J.; Jiang, S.; Lehrer, R.I.; Blumenthal, R. Theta-defensins prevent HIV-1 Env-mediated fusion by binding gp41 and blocking 6-helix bundle formation. *J. Biol. Chem.*, **2006**, *281*, 18787-18792.

[7] Wang, W.; Mulakala, C.; Ward, S.C.; Jung, G.; Luong, H.; Pham, D.; Waring, A.J.; Kaznessis, Y.; Lu, W.; Bradley, K.A.; Lehrer, R.I. Retrocyclins kill bacilli and germinating spores of Bacillus anthracis and inactivate anthrax lethal toxin. *J. Biol. Chem.*, **2006**, *281*, 32755-32764.

[8] Seidel, A.; Ye, Y.; de Armas, L.R.; Soto, M.; Yarosh, W.; Marcsisin, R.A.; Tran, D.; Selsted, M.E.; Camerini, D. Cyclic and acyclic defensins inhibit human immunodeficiency virus type-1 replication by different mechanisms. *PLoS One*, **2010**, *5(3)*, e9737.

[9] Craik, D.J. Circling the enemy: cyclic proteins in plant defence. *Trends Plant Sci.*, **2009**, *14(6)*, 328-335.

[10] Korsinczky, M.L.; Schirra, H.J.; Craik, D.J. Sunflower trypsin inhibitor-1. *Curr. Protein Pept. Sci.*, **2004**, *5(5)*, 351-364.

[11] Wang, C.K.; Kaas, Q.; Chiche, L.; Craik, D.J. CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering. *Nucleic Acids Res.*, **2008**, *36*, D206-210.

[12] Jennings, C.; West, J.; Waine, C.; Craik, D.; Anderson, M. Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from Oldenlandia affinis. *Proc. Natl. Acad. Sci. USA*, **2001**, *98(19)*, 10614-10619.

[13] Jennings, C.V.; Rosengren, K.J.; Daly, N.L.; Plan, M.; Stevens, J.; Scanlon, M.J.; Waine, C.; Norman, D.G.; Anderson, M.A.; Craik, D.J. Isolation, solution structure, and insecticidal activity of kalata

B2, a circular protein with a twist: do Mobius strips exist in nature? *Biochemistry*, **2005**, *44*, 851-860.

[14] Gran, L. Oxytocic principles of Oldenlandia affinis. *Lloydia*, **1973**, *36(2)*, 174-178.

[15] Gustafson, K.R.; Sowder, R.C.; Henderson, L.E.; Parsons, I.C.; Kashman, Y.; Cardellina, J.H.; McMahon, J.B.; Buckheit, R.W.; Pannell, L.K.; Boyd, M.R. Circulin-A and Circulin-B: Novel HIV-inhibitory macrocyclic peptides from the tropical tree *Chassalia parvifolia*. *J. Am. Chem. Soc.*, **1994**, *116*, 9337-9338.

[16] Tam, J.P.; Lu, Y.A.; Yang, J.L.; Chiu, K.W. An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 8913-8918.

[17] Gran, L.; Sletten, K.; Skjeldal, L. Cyclic peptides from Oldenlandia affinis DC. Molecular and biological properties. *Chem. Biodivers.*, **2008**, *5*, 2014-2022.

[18] Lindholm, P.; Goransson, U.; Johansson, S.; Claeson, P.; Gullbo, J.; Larsson, R.; Bohlin, L.; Backlund, A. Cyclotides: a novel type of cytotoxic agents. *Mol. Cancer The.r*, **2002**, *1*, 365-369.

[19] Colgrave, M.L.; Kotze, A.C.; Ireland, D.C.; Wang, C.K.; Craik, D.J. The anthelmintic activity of the cyclotides: natural variants with enhanced activity. *Chembiochem*, **2008**, *9*, 1939-1945.

[20] Colgrave, M.L.; Kotze, A.C.; Huang, Y.H.; O'Grady, J.; Simonsen, S.M.; Craik, D.J. Cyclotides: natural, circular plant peptides that possess significant activity against gastrointestinal nematode parasites of sheep. *Biochemistry*, **2008**, *47*, 5581-5589.

[21] Greenwood, K.P.; Daly, N.L.; Brown, D.L.; Stow, J.L.; Craik, D.J. The cyclic cystine knot miniprotein MCoTI-II is internalized into cells by macropinocytosis. *Int J. Biochem. Cell Biol.*, **2007**, *39*, 2252-2264.

[22] Colgrave, M.L.; Craik, D.J. Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: the importance of the cyclic cystine knot. *Biochemistry*, **2004**, *43*, 5965-5975.

[23] Craik, D.J.; Cemazar, M.; Wang, C.K.; Daly, N.L. The cyclotide family of circular miniproteins: nature's combinatorial peptide template. *Biopolymers*, **2006**, *84*, 250-266.

[24] Clark, R.J.; Daly, N.L.; Craik, D.J. Structural plasticity of the cyclic-cystine-knot framework: implications for biological activity and drug design. *Biochem. J.*, **2006**, *394*, 85-93.

[25] Simonsen, S.M.; Sando, L.; Rosengren, K.J.; Wang, C.K.; Colgrave, M.L.; Daly, N.L.; Craik, D.J. Alanine scanning mutagenesis of the prototypic cyclotide reveals a cluster of residues essential for bioactivity. *J. Biol. Chem.*, **2008**, *283*, 9805-9813.

[26] Austin, J.; Wang, W.; Puttamadappa, S.; Shekhtman, A.; Camarero, J.A. Biosynthesis and biological screening of a genetically encoded library based on the cyclotide MCoTI-I. *Chembiochem*, **2009**, *10*, 2663-2670.

[27] Jagadish, K.; Camarero, J.A. Cyclotides, a promising molecular scaffold for peptide-based therapeutics. *Biopolymers*, **2010**, DOI: 10.1002/bip.21433.

[28] Gillon, A.D.; Saska, I.; Jennings, C.V.; Guarino, R.F.; Craik, D.J.; Anderson, M.A. Biosynthesis of circular proteins in plants. *Plant J.*, **2008**, *53*, 505-515.

[29] Saska, I.; Gillon, A.D.; Hatsugai, N.; Dietzgen, R.G.; Hara-Nishimura, I.; Anderson, M.A.; Craik, D.J. An asparaginyl endopeptidase mediates *in vivo* protein backbone cyclization. *J. Biol. Chem.*, **2007**, *282*, 29721-29728.

[30] Sheldon, P.S.; Keen, J.N.; Bowles, D.J. Post-translational peptide bond formation during concanavalin A processing *in vitro*. *Biochem. J.*, **1996**, *320(Pt. 3)*, 865-870.

[31] Plan, M.R.; Saska, I.; Cagauan, A.G.; Craik, D.J. Backbone cyclised peptides from plants show molluscicidal activity against the rice pest Pomacea canaliculata (golden apple snail). *J. Agric. Food Chem.*, **2008**, *56*, 5237-5241.

[32] Gran, L. On the effect of a polypeptide isolated from "Kalata-Kalata" (Oldenlandia affinis DC) on the oestrogen dominated uterus. *Acta Pharmacol. Toxicol. (Copenh)*, **1973**, *33(5)*, 400-408.

[33] Gran, L.; Sandberg, F.; Sletten, K. Oldenlandia affinis (R&S) DC. A plant containing uteroactive peptides used in African traditional medicine. *J. Ethnopharmacol.*, **2000**, *70*, 197-203.

[34] Saether, O.; Craik, D.J.; Campbell, I.D.; Sletten, K.;Juul, J.; Norman, D.G. Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide kalata B1. *Biochemistry*, **1995**, *34*, 4147-4158.

[35] Craik, D.J.; Daly, N.L.; Bond, T.; Waine, C. Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *J. Mol. Biol.*, **1999**, *294*, 1327-1336.

[36] Goransson, U.; Luijendijk, T.; Johansson, S.; Bohlin, L.; Claeson, P. Seven novel macrocyclic polypeptides from Viola arvensis. *J. Nat. Prod.*, **1999**, *62*, 283-286.

[37] Henriques, S.T.; Craik, D.J. Cyclotides as templates in drug design. *Drug Discov. Today*, **2010**, *15*, 57-64.

[38] Daly, N.L.; Gustafson, K.R.; Craik, D.J. The role of the cyclic peptide backbone in the anti-HIV activity of the cyclotide kalata B1. *FEBS Lett..*, **2004**, *574*, 69-72.

[39] Chen, B.; Colgrave, M.L.; Daly, N.L.; Rosengren, K.J.; Gustafson, K.R.; Craik, D.J. Isolation and characterization of novel cyclotides from Viola hederaceae: solution structure and anti-HIV activity of vhl-1, a leaf-specific expressed cyclotide. *J. Biol. Chem.*, **2005**, *280*, 22395-22405.

[40] Wang, C.K.; Colgrave, M.L.; Gustafson, K.R.; Ireland, D.C.; Goransson, U.; Craik, D.J. Anti-HIV cyclotides from the Chinese medicinal herb Viola yedoensis. *J. Nat. Prod.*, **2008**, *71*, 47-52.

[41] Kamimori, H.; Hall, K.; Craik, D.J.; Aguilar, M.I. Studies on the membrane interactions of the cyclotides kalata B1 and kalata B6 on model membrane systems by surface plasmon resonance. *Anal. Biochem.*, **2005**, *337*, 149-153.

[42] Shenkarev, Z.O.; Nadezhdin, K.D.; Sobol, V.A.; Sobol, A.G.; Skjeldal, L.; Arseniev, A.S. Conformation and mode of membrane interaction in cyclotides. Spatial structure of kalata B1 bound to a dodecylphosphocholine micelle. *FEBS J.*, **2006**, *273*, 2658-2672.

[43] Shenkarev, Z.O.; Nadezhdin, K.D.; Lyukmanova, E.N.; Sobol, V.A.; Skjeldal, L.; Arseniev, A.S. Divalent cation coordination and mode of membrane interaction in cyclotides: NMR spatial structure of ternary complex Kalata B7/Mn2+/DPC micelle. *J. Inorg. Biochem.*, **2008**, *102*, 1246-1256.

[44] Wang, C.K.; Hu, S.H.; Martin, J.L.; Sjogren, T.; Hajdu, J.; Bohlin, L.; Claeson, P.; Goransson, U.; Rosengren, K.J.; Tang, J.; Tan, N.H.; Craik, D.J. Combined X-ray and NMR analysis of the stability of the cyclotide cystine knot fold that underpins its insecticidal activity and potential use as a drug scaffold. *J. Biol. Chem.*, **2009**, *284*, 10672-10683.

[45] Svangard, E.; Goransson, U.; Hocaoglu, Z.; Gullbo, J.; Larsson, R.; Claeson, P.; Bohlin, L. Cytotoxic cyclotides from Viola tricolor. *J. Nat. Prod.*, **2004**, *67*, 144-147.

[46] Herrmann, A.; Burman, R.; Mylne, J.S.; Karlsson, G.; Gullbo, J.; Craik, D.J.; Clark, R.J.; Goransson, U. The alpine violet, Viola biflora, is a rich source of cyclotides with potent cytotoxicity. *Phytochemistry*, **2008**, *69*, 939-952.

[47] Herrmann, A.; Svangard, E.; Claeson, P.; Gullbo, J.; Bohlin, L.; Goransson, U. Key role of glutamic acid for the cytotoxic activity of the cyclotide cycloviolacin O2. *Cell. Mol. Life Sci.*, **2006**, *63*, 235-245.

[48] Chen, B.; Colgrave, M.L.; Wang, C.; Craik, D.J. Cycloviolacin H4, a hydrophobic cyclotide from Viola hederaceae. *J. Nat. Prod.*, **2006**, *69*, 23-28.

[49] Huang, Y.H.; Colgrave, M.L.; Clark, R.J.; Kotze, A.C.; Craik, D.J. Lysine-scanning mutagenesis reveals an amendable face of the cyclotide kalata B1 for the optimization of nematocidal activity. *J. Biol. Chem.*, **2010**, *285*, 10797-10805.

[50] Witherup, K.M.; Bogusky, M.J.; Anderson, P.S.; Ramjit, H.; Ransom, R.W.; Wood, T.; Sardana, M. Cyclopsychotride A, a biologically active, 31-residue cyclic peptide isolated from Psychotria longipes. *J. Nat. Prod.*, **1994**, *57*, 1619-1625.

[51] Marglin, A.; Merrifield, R.B. Chemical synthesis of peptides and proteins. *Annu. Rev. Biochem.*, **1970**, *39*, 841-866.

[52] Tam, J.P.; Lu, Y.-A. Synthesis of Large Cyclic Cystine-Knot Peptide by Orthogonal Coupling Strategy Using Unprotected Peptide Precursor. *Tetrathedron Lett.*, **1997**, *38*, 5599-5602.

[53] Tam, J.P.; Lu, Y.A. A biomimetic strategy in the synthesis and fragmentation of cyclic protein. *Protein Sci..*, **1998**, *7*, 1583-1592.

[54] Daly, N.L.; Love, S.; Alewood, P.F.; Craik, D.J. Chemical synthesis and folding pathways of large cyclic polypeptides: studies of the cystine knot polypeptide kalata B1. *Biochemistry*, **1999**, *38*, 10606-10614.

[55] Thongyoo, P.; Tate, E.W.; Leatherbarrow, R.J. Total synthesis of the macrocyclic cysteine knot microprotein MCoTI-II. *Chem. Commun. (Camb.)*, **2006**, 2848-2850.

[56] Leta Aboye, T.; Clark, R.J.; Craik, D.J.; Goransson, U. Ultra-stable peptide scaffolds for protein engineering-synthesis and folding of the circular cystine knotted cyclotide cycloviolacin O2. *Chembiochem*, **2008**, *9*, 103-113.

[57] Dawson, P.E.; Muir, T.W.; Clark-Lewis, I.; Kent, S.B.H. Synthesis of Proteins by Native Chemical Ligation. *Science*, **1994**, *266*, 776-779.

[58] Camarero, J.A.; Muir, T.W. Chemoselective backbone cyclization of unprotected peptides. *Chem. Commun*., **1997**, *1997*, 1369-1370.

[59] Camarero, J.A.; Muir, T.W. Biosynthesis of a Head-to-Tail Cyclized Protein with Improved Biological Activity. *J. Am. Chem. Soc*., **1999**, *121*, 5597-5598.

[60] Camarero, J.A.; Mitchell, A.R. Synthesis of proteins by native chemical ligation using Fmoc-based chemistry. *Protein Pept. Lett.*, **2005**, *12*, 723-728.

[61] Camarero, J.A.; Cotton, G.J.; Adeva, A.; Muir, T.W. Chemical ligation of unprotected peptides directly from a solid support. *J. Pept. Res.*, **1998**, *51*, 303-316.

[62] Beligere, G.S.; Dawson, P.E. Conformationally assisted ligation using C-terminal thioester peptides. *J. Am. Chem. Soc*., **1999**, *121*, 6332-6333.

[63] Camarero, J.A.; Adeva, A.; Muir, T.W. 3-Thiopropionic acid as a highly versatile multidetachable thioester resin linker. *Lett. Pept. Sci.*, **2000**, *7*, 17-21.

[64] Ingenito, R.; Bianchi, E.; Fattori, D.; Pessi, A. Solid phase synthesis of peptide C-terminal thioesters by Fmoc/t-Bu chemistry Source. *J. Am. Chem. Soc.*, **1999**, *121*, 11369-11374.

[65] Shin, Y.; Winans, K.A.; Backes, B.J.; Kent, S.B.H.; Ellman, J.A.; Bertozzi, C.R. Fmoc-Based Synthesis of Peptide-aThioesters: Application to the total chemical synthesis of a glycoprotein by native chemical ligation. *J. Am. Chem. Soc.*, **1999**, *121*, 11684-11689.

[66] Camarero, J.A.; Hackel, B.J.; de Yoreo, J.J.; Mitchell, A.R. Fmoc-based synthesis of peptide alpha-thioesters using an aryl hydrazine support. *J. Org. Chem*., **2004**, *69*, 4145-4151.

[67] Blanco-Canosa, J.B.; Dawson, P.E. An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. *Angew. Chem. Int. Ed. Engl*., **2008**, *47*, 6851-6855.

[68] Sancheti, H.; Camarero, J.A. "Splicing up" drug discovery. Cell-based expression and screening of genetically-encoded libraries of backbone-cyclized polypeptides. *Adv. Drug Deliv. Rev*., **2009**, *61*, 908-917.

[69] Kimura, R.H.; Tran, A.T.; Camarero, J.A. Biosynthesis of the cyclotide kalata B1 by using protein splicing. *Angew. Chem. Int. Ed. Engl*., **2006**, *45*, 973-976.

[70] Camarero, J.A.; Kimura, R.H.; Woo, Y.H.; Shekhtman, A.; Cantor, J. Biosynthesis of a fully functional cyclotide inside living bacterial cells. *Chembiochem*, **2007**, *8*, 1363-1366.

[71] Perler, F.B.; Adam, E. Protein splicing and its applications. *Curr. Opin. Biotechnol*., **2000**, *11*, 377-383.

[72] Hirel, P.H.; Schmitter, M.J.; Dessen, P.; Fayat, G.; Blanquet, S. Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl. Acad. Sci. USA*, **1989**, *86*, 8247-8251.

[73] Dwyer, M.A.; Lu, W.; Dwyer, J.J.; Kossiakoff, A.A. Biosynthetic phage display: a novel protein engineering tool combining chemical and genetic diversity. *Chem. Biol.*, **2000**, *7*, 263-274.

[74] Iwai, H.; Pluckthum, A. Circular b-lactamase: stability enhancement by cyclizing the backbone. *FEBS Lett*., **1999**, *459*, 166-172.

[75] Camarero, J.A.; Fushman, D.; Cowburn, D.; Muir, T.W. Peptide chemical ligation inside living cells: *in vivo* generation of a circular protein domain. *Bioorg. Med. Chem*., **2001**, *9*, 2479-2484.

[76] Cotton, G.J.; Ayers, B.; Xu, R.; Muir, T.W. Insertion of a Synthetic Peptide into a Recombinant Protein Framework; A Protein Biosensor. *J. Am. Chem. Soc*., **1999**, *121*, 1100-1101.

[77] Tolbert, T.J.; Wong, C.-H. New methods for proteomic research: preparation of proteins with N-terminal cysteines for labeling and conjugation. *Angew. Chem. Int. Ed*., **2002**, *41*, 2171-2174.

[78] Erlanson, D.A.; Chytil, M.; Verdine, G.L. The leucine zipper domain controls the orientation of AP-1 in the NFAT.AP-1.DNA complex. *Chem. Biol*., **1996**, *3*, 981-991.

[79] Baker, R.T.; Smith, S.A.; Marano, R.; McKee, J.; Board, P.G. Protein expression using cotranslational fusion and cleavage of ubiquitin. Mutagenesis of the glutathione-binding site of human Pi class glutathione S-transferase. *J. Biol. Chem.*, **1994**, *269*, 25381-25386.

[80] Varshavsky, A. Ubiquitin fusion technique and related methods. *Methods Enzymol.*, **2005**, *399*, 777-799.

[81] Dykes, C.W.; Bookless, A.B.; Coomber, B.A.; Noble, S.A.; Humber, D.C.; Hobden, A.N. Expression of atrial natriuretic factor as a cleavable fusion protein with chloramphenicol acetyltransferase in Escherichia coli. *Eur. J. Biochem*., **1988**, *174*, 411-416.

[82] Liu, D.; Xu, R.; Dutta, K.; Cowburn, D. N-terminal cysteinyl proteins can be prepared using thrombin cleavage. *FEBS Lett.*, **2008**, *582*, 1163-1167.

[83] Dalbey, R.E.; Lively, M.O.; Bron, S.; van Dijl, J.M. The chemistry and enzymology of the type I signal peptidases. *Protein Sci*., **1997**, *6*, 1129-1138.

[84] Paetzel, M.; Dalbey, R.E.; Strynadka, N.C. Crystal structure of a bacterial signal peptidase apoenzyme: implications for signal peptide binding and the Ser-Lys dyad mechanism. *J. Biol. Chem*., **2002**, *277*, 9512-9519.

[85] Hauser, P.S.; Ryan, R.O. Expressed protein ligation using an N-terminal cysteine containing fragment generated *in vivo* from a pelB fusion protein. *Protein Expr. Purif*., **2007**, *54*, 227-233.

[86] Evans, T.C.; Benner, J.; Xu, M.-Q. The *in Vitro* Ligation of Bacterially Expressed Proteins Using an Intein from *Metanobacterium thermoautotrophicum*. *J. Biol. Chem*., **1999**, *274*, 3923-3926.

[87] Southworth, M.W.; Amaya, K.; Evans, T.C.; Xu, M.Q.; Perler, F.B. Purification of proteins fused to either the amino or carboxy terminus of the Mycobacterium xenopi gyrase A intein. *Biotechniques*, **1999**, *27*, 110-114, 116, 118-120.

[88] Mathys, S.; Evans, T.C.; Chute, I.C.; Wu, H.;Chong, S.; Benner, J.; Liu, X.Q.; Xu, M.Q. Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. *Gene*, **1999**, *231*, 1-13.

[89] Austin, J.; Kimura, R.H.; Woo, Y.H.; Camarero, J.A. *In vivo* biosynthesis of an Ala-scan library based on the cyclic peptide SFTI-1. *Amino Acids*, **2010**, *38*, 1313-1322.

[90] Abel-Santos, E.; Scott, C.P.; Benkovic, S.J. Use of inteins for the *in vivo* production of stable cyclic peptide libraries in E. coli. *Methods Mol. Biol.*, **2003**, *205*, 281-294.

[91] Tavassoli, A.; Benkovic, S.J. Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in E. coli. *Nat. Protoc*., **2007**, *2*, 1126-1133.

[92] Scott, C.P.; Abel-Santos, E.; Wall, M.; Wahnon, D.; Benkovic, S.J. Production of cyclic peptides and proteins *in vivo*. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 13638-13643.

[93] Scott, C.P.; Abel-Santos, E.; Jones, A.D.; Benkovic, S.J. Structural requirements for the biosynthesis of backbone cyclic peptide libraries. *Chem. Biol*., **2001**, *8*, 801-815.

[94] Gunasekera, S.; Foley, F.M.; Clark, R.J.; Sando, L.; Fabri, L.J.; Craik, D.J.; Daly, N.L. Engineering stabilized vascular endothelial growth factor-A antagonists: synthesis, structural characterization, and bioactivity of grafted analogues of cyclotides. *J. Med. Chem*., **2008**, *51*, 7697-7704.

[95] Thongyoo, P.; Roque-Rosell, N.; Leatherbarrow, R.J.; Tate, E.W. Chemical and biomimetic total syntheses of natural and engineered MCoTI cyclotides. *Org. Biomol. Chem.*, **2008**, *6*, 1462-1470.

[96] Thongyoo, P.; Bonomelli, C.; Leatherbarrow, R.J.; Tate, E.W. Potent inhibitors of beta-tryptase and human leukocyte elastase based on the MCoTI-II scaffold. *J. Med. Chem*., **2009**, *52*, 6197-6200.

[97] Reiss, S.; Sieber, M.; Oberle, V.; Wentzel, A.; Spangenberg, P.; Claus, R.; Kolmar, H.; Losche, W. Inhibition of platelet aggregation by grafting RGD and KGD sequences on the structural scaffold of small disulfide-rich proteins. *Platelets*, **2006**, *17*, 153-157.

[98] Gruber, C.W.; Muttenthaler, M.; Freissmuth, M. Ligand-based peptide design and combinatorial peptide libraries to target G protein-coupled receptors. *Curr. Pharm. Des*., **2010**, [Epub ahead of print].

[99] Kinsella, T.M.; Ohashi, C.T.; Harder, A.G.; Yam, G.C.; Li, W.;Peelle, B.; Pali, E.S.; Bennett, M.K.; Molineaux, S.M.; Anderson, D.A.; Masuda, E.S.; Payan, D.G. Retrovirally delivered random cyclic Peptide libraries yield inhibitors of interleukin-4 signaling in human B cells. *J. Biol. Chem.*, **2002**, *277*, 37512-37518.

[100] Cheng, L.; Naumann, T.A.; Horswill, A.R.; Hong, S.J.; Venters, B.J.; Tomsho, J.W.; Benkovic, S.J.; Keiler, K.C. Discovery of antibacterial cyclic peptides that inhibit the ClpXP protease. *Protein Sci.*, **2007**, *16*, 1535-1542.

[101] Tavassoli, A.; Lu, Q.; Gam, J.; Pan, H.; Benkovic, S.J.; Cohen, S.N. Inhibition of HIV budding by a genetically selected cyclic peptide targeting the Gag-TSG101 interaction. *ACS Chem. Biol.*, **2008**, *3*, 757-764.

[102] Hernandez, J.F.; Gagnon, J.; Chiche, L.;Nguyen, T.M.; Andrieu, J.P.; Heitz, A.; Trinh Hong, T.; Pham, T.T.; Le Nguyen, D. Squash trypsin inhibitors from Momordica cochinchinensis exhibit an atypical macrocyclic structure. *Biochemistry*, **2000**, *39*, 5722-5730.

[103] Giepmans, B.N.; Adams, S.R.; Ellisman, M.H.; Tsien, R.Y. The fluorescent toolbox for assessing protein location and function. *Science*, **2006**, *312*, 217-224.

[104] You, X.; Nguyen, A.W.; Jabaiah, A.; Sheff, M.A.; Thorn, K.S.; Daugherty, P.S. Intracellular protein interaction mapping with FRET hybrids. *Proc. Natl. Acad. Sci. USA*, **2006**, *103*, 18458-18463.

[105] Daly, N.L.; Rosengren, K.J.; Craik, D.J. Discovery, structure and biological activities of cyclotides. *Adv. Drug Deliv. Rev.*, **2009**, *61*, 918-930.

[106] Kimura, R.H.; Steenblock, E.R.; Camarero, J.A. Development of a cell-based fluorescence resonance energy transfer reporter for Bacillus anthracis lethal factor protease. *Anal. Biochem.*, **2007**, *369*, 60-70.

[107] Kolmar, H. Biological diversity and therapeutic potential of natural and engineered cystine knot miniproteins. *Curr. Opin. Pharmacol.*, **2009**, *9*, 608-614.

[108] Rosengren, K.J.; Daly, N.L.; Plan, M.R.; Waine, C.; Craik, D.J. Twists, knots, and rings in proteins. Structural definition of the cyclotide framework. *J. Biol. Chem.*, **2003**, *278*, 8606-8616.

[109] Felizmenio-Quimio, M.E.; Daly, N.L.; Craik, D.J. Circular proteins in plants: solution structure of a novel macrocyclic trypsin inhibitor from Momordica cochinchinensis. *J. Biol. Chem.*, **2001**, *276*, 22875-22882.

Received: July 07, 2010 Revised: August 18, 2010 Accepted: August 25, 2010

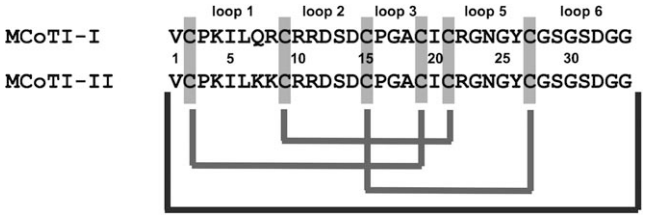
Molecular Dynamics

DOI: 10.1002/anie.201002906

Backbone Dynamics of Cyclotide MCoTI-I Free and Complexed with Trypsin**

*Shadakshara S. Puttamadappa, Krishnappa Jagadish, Alexander Shekhtman, and Julio A. Camarero**

Cyclotides are a new emerging family of large plant-derived backbone-cyclized polypeptides (about 28–37 amino acids long) that share a disulfide-stabilized core (three disulfide bonds) characterized by an unusual knotted arrangement.[1] Cyclotides contrast with other circular polypeptides in that they have a well-defined three-dimensional structure, and despite their small size can be considered as microproteins. Their unique circular backbone topology and knotted arrangement of three disulfide bonds makes them exceptionally stable to thermal and enzymatic degradation (Scheme 1).



Scheme 1. Primary structure and disulfide connectivities of MCoTI cyclotides. Dark gray and light gray connectors represent peptide and disulfide bonds, respectively.

Furthermore, their well-defined structures have been associated with a wide range of biological functions.[2,3] Cyclotides MCoTI-I/II are powerful trypsin inhibitors ($K_i \approx 20$–30 pM) that have been recently isolated from the dormant seeds of *Momordica cochinchinensis*, a plant member of the cucurbitaceae family.[4] Although MCoTI cyclotides do not share significant sequence homology with other cyclotides beyond the presence of the three cystine bridges, structural analysis by NMR spectroscopy has shown that they adopt a similar backbone-cyclic cystine-knot topology.[5,6] MCoTI cyclotides, however, show high sequence homology with related cystine-knot squash trypsin inhibitors,[4] and therefore represent interesting molecular scaffolds for drug design.[7–10]

Determination of the backbone dynamics of these fascinating microproteins is key for understanding their physical and biological properties. Internal motions of a protein on different timescales, extending from picoseconds to a second, have been suggested to play an important role in its biological function.[11] A better understanding of the backbone dynamics of the cyclotide scaffold will be extremely helpful for evaluating its utility as a scaffold for peptide-based drug discovery. Such insight will help in the design of optimal focused libraries that can be used for the discovery of new cyclotide sequences with novel biological activities.[12,13]

Herein, we report for the first time the determination of the internal dynamics of the cyclotide MCoTI-I in the free state and complexed with trypsin. Uniformly ^{15}N-labeled natively folded cyclotide MCoTI-I was recombinantly produced in *Escherichia coli* growing in minimal M9 medium containing $^{15}NH_4Cl$ as the only source of nitrogen. Concomitant backbone cyclization and folding were accomplished by using intramolecular native chemical ligation[14,15] in combination with a modified protein splicing unit (Figure S1, Supporting Information).[16–18] The internal dynamics of cyclotide MCoTI-I was obtained from ^{15}N spin–lattice and spin–spin relaxation times and $^{15}N\{^1H\}$ heteronuclear Overhauser effect (NOE) enhancements.[11] The backbone flexibility was characterized by the square of the generalized order parameter, S^2, which reveals the dynamics of backbone NH groups on the pico- to nanosecond timescale.[19,20] The order parameter satisfies the inequality $0 \le S^2 \le 1$, in which lower values indicate larger amplitudes of intramolecular motions. Motions on the milli- to microsecond timescale were assessed by the presence of the chemical exchange terms in the spin–spin relaxation.

The NMR spectrum and S^2 values, derived from the ^{15}N relaxation data of free MCoTI-I, are shown in Figure 1 a and d, respectively. Residues Ile5 and Gly23 of free MCoTI-I were excluded from the backbone dynamics analysis since the relaxation data could not be fitted to a monoexponential function, possibly as a result of chemical exchange.[21] Gln7 was not assigned because of broadening of the NMR signal, presumably caused by fast exchange with water. The S^2 values for free MCoTI-I show that most of the NH groups of the cyclotide backbone are highly constrained with S^2 values ≥ 0.8, thus resembling those found in well-folded globular proteins (Table 1). The average S^2 value, $\langle S^2 \rangle$, for free MCoTI-I was 0.83 ± 0.03. This value is similar to that found

[*] Dr. K. Jagadish, Dr. J. A. Camarero
Department of Pharmaceutical Sciences and Pharmacology
University of Southern California
Los Angeles, CA 90033 (USA)
Fax: (+1) 323-224-7473
E-mail: jcamarer@usc.edu
S. S. Puttamadappa, Dr. A. Shekhtman
Department of Chemistry, State University of New York
Albany, NY 12222 (USA)

[**] This work was supported by funding from the School of Pharmacy at the University of Southern California, by National Institute of Health award GM090323-01 to J.A.C., and by American Diabetes Association award 1-06-CD-23 and National Institute of Health award GM090323-01 to A.S.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201002906.

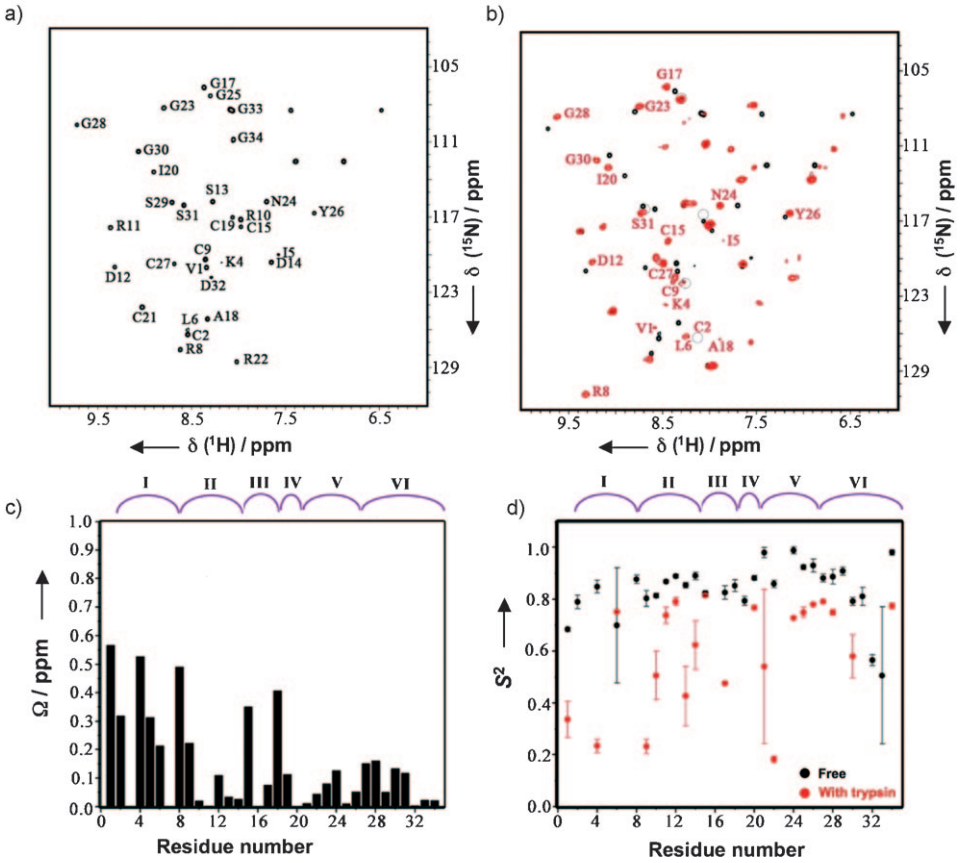


Figure 1. NMR analysis of the backbone dynamics of free and trypsin-bound MCoTI-I. a) $\{^{15}N,^{1}H\}$ NMR heteronuclear single quantum correlation (HSQC) spectrum of free MCoTI-I. Chemical shift assignments of the backbone amides are indicated. b) Overlay of the $\{^{15}N,^{1}H\}$ HSQC spectra of free (black) and trypsin-bound MCoTI-I (red). Residues with large average amide chemical shift differences between two different states (>0.3 ppm) are indicated. Peaks that are broadened in trypsin-bound MCoTI-I are indicated by gray circles. c) Average amide chemical shift difference for all the assigned residues in free and trypsin-bound MCoTI-I. The chemical shift difference was calculated as: $\Delta\Omega = [(\Delta\Omega^2_{NH} + 0.04\,\Omega^2_N)/2]^{1/2}$, where $\Delta\Omega_{NH}$ and $\Delta\Omega_N$ are the changes in the amide proton and nitrogen chemical shifts (ppm), respectively. d) Order parameter, S^2, for free (black) and trypsin-bound MCoTI-I (red). The S^2 value is a measure of backbone flexibility and represents the degree of angular restriction of the N–H vector in the molecular frame. The MCoTI-I loops are shown at the top of (c) and (d). Small unassigned peaks in the spectra of both free and trypsin-bound MCoTI-I are from a minor conformation of the protein, and result from a known isomerization of the backbone at an Asp–Gly sequence in loop 6 of MCoTI-I.

Table 1: Average order parameters of structural elements in MCoTI-I in the free state and bound to trypsin.

Structural element	Sequence	$<S^2>$ [a] Free MCoTI-I	$<S^2>$ [b] Trypsin–MCoTI-I
loop 1	3–8	0.81 ± 0.01	0.49 ± 0.05
loop 2	10–14	0.81 ± 0.01	0.62 ± 0.07
loop 3	16–18	0.84 ± 0.02	0.48[c]
loop 4	20	0.88[c]	0.76[c]
loop 5	22–26	0.92 ± 0.02	0.61 ± 0.01
loop 6	28–34	0.76 ± 0.05	0.61 ± 0.05
cystine knot	2,10,15,19,21,27	0.84 ± 0.02	0.60 ± 0.08

[a] S^2 values for residues 5 and 23 from free MCoTI-I are not included in the average because the relaxation data could not be fitted to a monoexponential function. [b] S^2 values for residues 2, 5, 8, 18, 19, 23, 29, 31, 32, and 33 from trypsin-bound MCoTI-I are not included in the average because of the lack of signal intensity or because the relaxation data could not be fitted to a monoexponential function. [c] $<S^2>$ contains the S^2 value for a single residue.

for the six cystine residues involved in the cystine knot ($<S^2> = 0.84 \pm 0.02$) and is considerably larger than those found for other linear squash trypsin inhibitors ($<S^2> = 0.71$ for trypsin inhibitor from *Cucurbita maxima* (CMTI-III, 78% homology with MCoTI-I)),[22] thus indicating the importance of the backbone cyclization to rigidifying the overall structure. Loops 2 through 5 in free MCoTI-I showed $<S^2>$ values ≥ 0.8. In particular, loop 5 showed an $<S^2>$ value of 0.92 ± 0.02, well above the average for the molecule and the cystine knot. In contrast, loops 1 and 6 showed $<S^2>$ values below the average for the molecule. Thus, loop 6, which is believed to act as a very flexible linker to allow cyclization,[23] had an $<S^2>$ value of 0.76 ± 0.17 with only two residues, Asp32 and Gly33, having values below 0.6. Despite this small $<S^2>$ value, residues in loop 6 did not require significant chemical exchange terms (Figure S2 and Table S1, Supporting Information), which suggests that the mobility observed arises mostly from local vibrations.

The $<S^2>$ value for loop 1, which is responsible for binding trypsin, was 0.81 ± 0.07. This value is $\approx 90\%$ of the average value for free MCoTI-I. Residue Leu6 in loop 1 also required chemical exchange terms to be considered, thus indicating the existence of intramolecular conformational exchange on the micro- to millisecond timescale. The mobility observed in loop 1 at both nano- to picosecond and millisecond timescales has also been described in other trypsin inhibitors,[22,24,25] and it has been suggested to play an important role in receptor–ligand binding.[11]

To explore whether that was the case in the MCoTI cyclotides, we next studied the effect of ligand binding on the backbone dynamics of MCoTI-I (Figures 1 and 2). To exclude the possibility that trypsin could cleave or scramble the disulfide bonds of MCoTI-I upon complex formation, we used a competition experiment of trypsin–[^{15}N]MCoTI-I with unlabeled MCoTI-I. The results indicated that the structure of MCoTI-I is unaltered upon trypsin binding (Figure S3, Supporting Information). Trypsin binding led to large (>0.3 ppm) and specific changes in the chemical shifts of

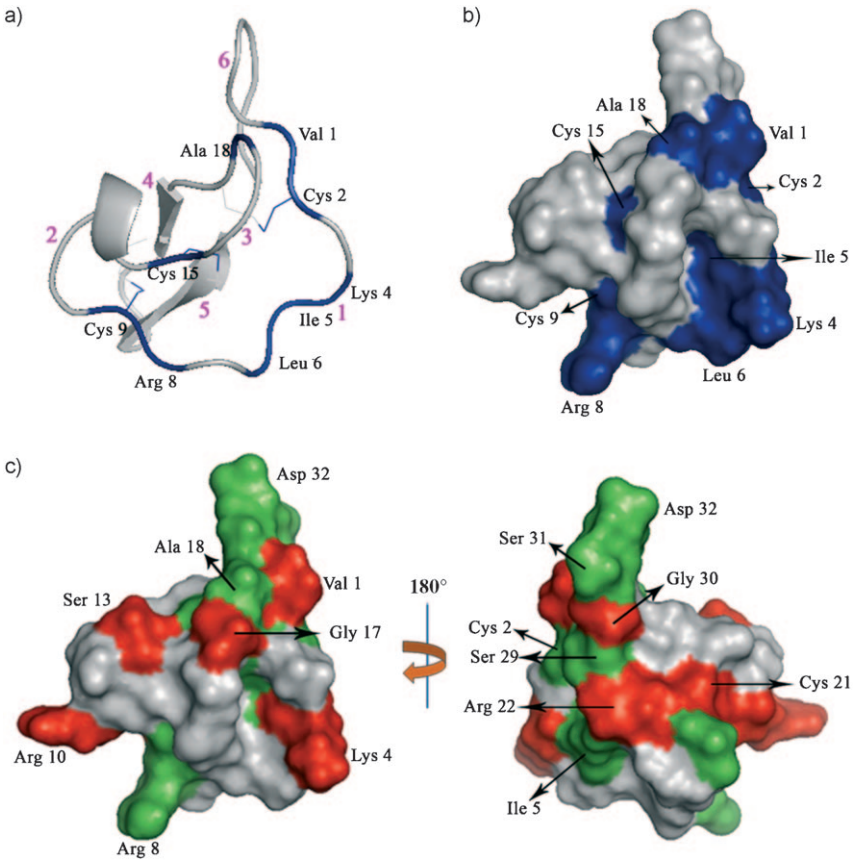


Figure 2. Trypsin binding to MCoTI-I affects the MCoTI-I backbone dynamics. a) Ribbon and b) surface diagrams of the trypsin–MCoTI-I interaction map. Red numbers indicate the positions of the MCoTI-I loops. The MCoTI-I residues with a large chemical shift difference (>0.3 ppm) are in blue. c) Changes in the MCoTI-I order parameter as a result of binding to trypsin. Residues with $S_f^2 - S_b^2 > 0.2$, where S_f^2 and S_b^2 are the order parameters of the free and trypsin-bound MCoTI-I, respectively, are depicted in red. MCoTI-I residues that were broadened in $\{^{15}N,^1H\}$ HSQC because of binding to trypsin are shown in green. The structure of free MCoTI/II (PDB code: 1IB9)[6] was used to illustrate the changes of MCoTI-I dynamics arising from trypsin binding.

the residues located in loop 1 (Cys2, Lys4, Ile5, Arg8), loop 3 (Cys15 and Ala18), and loop 6 (Val1) (Figures 1c and 2b and Table S2, Supporting Information). NMR signals of Cys2, Ile5, Cys19, Ser29, Ser31, Asp32, and Gly33 were significantly broadened, presumably because of intramolecular chemical exchanges in the trypsin–MCoTI-I complex. Arg8, Ala18, and Gly23 were excluded from backbone dynamics analysis because their peaks were broadened in the $^{15}N\{^1H\}$ NOE spectra. Similar findings have already been reported for other biomolecular interactions.[26]

We used these changes to construct the trypsin–MCoTI-I interaction surface. The binding surface is contiguous and spans 46% of the total molecular area of MCoTI-I (Figure 2b). As expected, the major difference in the backbone dynamics was observed in the binding loop (Table 1), where the mobility in the nano- to picosecond timescale was increased in MCoTI once bound to trypsin. Loop 1 showed $< S^2 > = 0.49 \pm 0.02$, which is much lower than the value for the rest of the molecule ($< S^2 > = 0.65 \pm 0.07$). Several residues in loop 2 (Cys9, Arg10, Ser13, and Asp14), loop 3 (Gly17), and loop 5 (Cys21 and Arg22) also showed significantly lower values of S^2 upon complex formation (Figures 1d and 2c). It is likely that the increase in mobility observed in these loops may help to accommodate the increased flexibility of the binding loop (Figure 2c).

Since our data clearly show that backbone flexibility of cyclotide MCoTI-I increases significantly upon binding to trypsin, we decided to estimate the contribution of these motions to the overall Gibbs free energy of binding (ΔG). The energetic benefit of this increase in backbone flexibility can be estimated from the experimental relaxation data, by using the experimentally measured order parameters S^2.[27] The estimated ΔG value was approximately $-62\ kJ\,mol^{-1}$ at 298 K. This value is almost identical to the calculated value from the trypsin inhibitory constant of MCoTI-I ($K_i \approx 20$ pM,[28] $\Delta G \approx -61\ kJ\,mol^{-1}$). The calculated entropic contribution ($-T\Delta S$) at the same temperature was approximately $-46\ kJ\,mol^{-1}$. These results highlight the importance of the backbone entropic term to the formation of the trypsin–MCoTI-I complex, although a more detailed thermodynamic analysis that also includes the side-chain motions may be required.

In summary, we have reported the backbone dynamics of the cyclotide MCoTI-I in the free state and complexed to its binding partner trypsin in solution. To our knowledge this is the first time the backbone dynamics of a natively folded cyclotide has been reported. This has been possible because of the use of modified protein splicing units for the heterologous expression of folded cyclotides using bacterial expression systems[17,18,29] to incorporate NMR-active nuclei such as ^{15}N. Our results on the backbone dynamics of free cyclotide MCoTI-I confirm that MCoTI-I adopts a well-folded and highly compact structure with an $< S^2 >$ value of 0.83. This value is similar to those found in the regions of well-folded proteins with restricted backbone dynamics.

The results also indicate that the trypsin-binding loop (loop 1) has a smaller S^2 value than the average value for the whole molecule, thus indicating a higher mobility of this region in the pico- to nanosecond timescale. This region also showed significant conformational exchange motions in the micro- to millisecond timescale. Loop 6 also possesses a higher mobility in the pico- to nanosecond timescale than the averaged value for MCoTI-I, although no significant conformational exchange motions were detected in the micro- to millisecond timescale. This result is intriguing since this loop contains a potentially flexible Gly–Ser-rich sequence that is mostly absent among other linear trypsin squash inhibitors,

and therefore it was thought to be a highly flexible linker to allow cyclization. More surprising, however, was the fact that the backbone of MCoTI-I, and especially loop 1, increased the pico- to nanosecond mobility when bound to trypsin. This interesting result has already been observed in other high-affinity protein–protein interactions.[30,31]

The thermodynamic analysis of the backbone contribution to the formation of the trypsin–MCoTI-I complex by using measured S^2 values also revealed the importance of the backbone entropic term in the formation of the complex. Similar findings have also been found in other protease inhibitors.[32] This increment in backbone mobility may help to minimize the entropic penalties required for binding. Hence, we also observed in the HSQC spectrum of the trypsin–MCoTI-I complex the appearance of a signal corresponding to the ε-NH_3^+ of Lys4 located in loop 1, which suggests that the ammonium group is protected and more rigid when forming the complex (Figure 3). This is the only

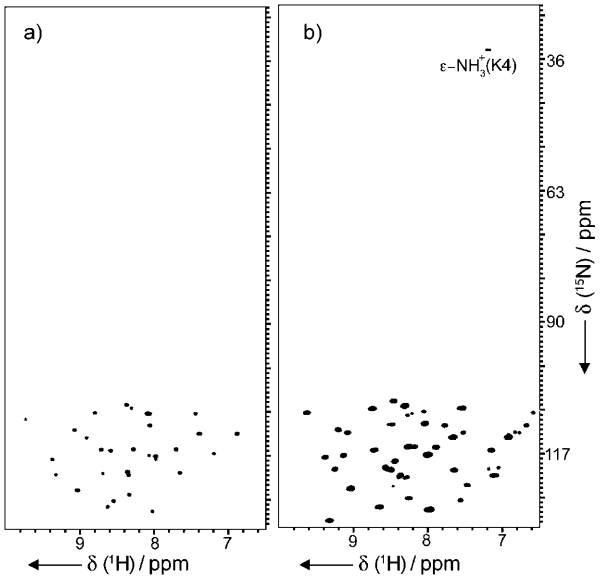


Figure 3. ε-NH_3^+ of Lys4 is protected from fast exchange with the solvent in trypsin-bound MCoTI-I. {^{15}N,^{1}H}-HSQC spectra of free (a) and trypsin-bound MCoTI-I (b) were collected at room temperature with the ^{15}N-carrier position at 82 ppm and ^{15}N radio-frequency field strengths of 5.2 kHz for 90° and 180° pulses and 1.2 kHz for composite decoupling during acquisition.

Lys residue present in the sequence of MCoTI-I (Scheme 1) and therefore it can be unambiguously assigned. This residue is key for binding to trypsin[29] and is responsible for binding to the specificity pocket of trypsin. This cross-peak was totally absent in the free MCoTI-I sample, which indicates that the ε-NH_3^+ of Lys4 is less rigid and rapidly exchanging with solvent (Figure 3a).

Similarly, the broadening of aliphatic resonances for Arg side chains with essentially rigid guanidinium groups (that is, $^{\varepsilon}$N–H bond vectors) has also been described for protein–peptide complexes.[26] Palmer and co-workers have recently suggested that this dynamic decoupling between the side-chain terminus from the rest of the aliphatic part of the side chain may be a general biophysical strategy for maximizing residual side-chain and potentially backbone conformational entropy in proteins and their complexes,[33] which is in agreement with our observations regarding the increase in MCoTI-I backbone mobility upon complex formation.

We have also mapped the binding surface of MCoTI-I once bound to trypsin. Major changes in chemical shifts were observed for the solvent-exposed residues located in loops 1 (Lys4, Ile5, Arg8), 3 (Ala18), and 6 (Val1) (Figures 1c and 2b). In agreement with these results, we have recently shown that the introduction of nonconservative mutations in these positions has a negative effect on the affinity for trypsin,[29] thus indicating that they may be in close contact with the protease at the binding interface of the molecular complex.

Cyclotides present several characteristics that make them appear as promising leads or frameworks for peptide drug design.[7,8] Investigation of the backbone dynamics is crucial for a better understanding of the dynamic structural properties of the cyclotide scaffold and how it affects the mode of binding of these interesting molecules. The reported data will help in the design of cyclotide-based libraries for molecular screening and the selection of de novo sequences with new biological activities, or the development of grafted analogues for use as peptide-based drugs.[9,10]

Received: May 14, 2010
Published online: August 16, 2010

Keywords: cyclotides · molecular dynamics · NMR spectroscopy · protein–protein interactions · structural biology

[1] N. L. Daly, K. J. Rosengren, D. J. Craik, *Adv. Drug Delivery Rev.* **2009**, *61*, 918.
[2] D. J. Craik, S. Simonsen, N. L. Daly, *Curr. Opin. Drug Discovery Dev.* **2002**, *5*, 251.
[3] D. J. Craik, M. Cemazar, C. K. Wang, N. L. Daly, *Biopolymers* **2006**, *84*, 250.
[4] J. F. Hernandez, J. Gagnon, L. Chiche, T. M. Nguyen, J. P. Andrieu, A. Heitz, T. Trinh Hong, T. T. Pham, D. Le Nguyen, *Biochemistry* **2000**, *39*, 5722.
[5] A. Heitz, J. F. Hernandez, J. Gagnon, T. Trinh Hong, T. T. Pham, T. M. Nguyen, D. Le-Nguyen, L. Chiche, *Biochemistry* **2001**, *40*, 7973.
[6] M. E. Felizmenio-Quimio, N. L. Daly, D. J. Craik, *J. Biol. Chem.* **2001**, *276*, 22875.
[7] R. J. Clark, N. L. Daly, D. J. Craik, *Biochem. J.* **2006**, *394*, 85.
[8] D. J. Craik, M. Cemazar, N. L. Daly, *Curr. Opin. Drug Discovery Dev.* **2006**, *9*, 251.
[9] D. J. Craik, N. L. Daly, J. Mulvenna, M. R. Plan, M. Trabi, *Curr. Protein Pept. Sci.* **2004**, *5*, 297.
[10] P. Thongyoo, N. Roque-Rosell, R. J. Leatherbarrow, E. W. Tate, *Org. Biomol. Chem.* **2008**, *6*, 1462.
[11] V. A. Jarymowycz, M. J. Stone, *Chem. Rev.* **2006**, *106*, 1624.
[12] H. Sancheti, J. A. Camarero, *Adv. Drug Delivery Rev.* **2009**, *61*, 908.
[13] D. C. Ireland, M. L. Colgrave, N. L. Daly, D. J. Craik, *Adv. Exp. Med. Biol.* **2009**, *611*, 477.
[14] J. A. Camarero, T. W. Muir, *Chem. Commun.* **1997**, 1369.
[15] J. A. Camarero, J. Pavel, T. W. Muir, *Angew. Chem.* **1998**, *110*, 361; *Angew. Chem. Int. Ed.* **1998**, *37*, 347.
[16] J. A. Camarero, T. W. Muir, *J. Am. Chem. Soc.* **1999**, *121*, 5597.

[17] R. H. Kimura, A. T. Tran, J. A. Camarero, *Angew. Chem.* **2006**, *118*, 987; *Angew. Chem. Int. Ed.* **2006**, *45*, 973.
[18] J. A. Camarero, R. H. Kimura, Y. H. Woo, A. Shekhtman, J. Cantor, *ChemBioChem* **2007**, *8*, 1363.
[19] G. Lipari, A. Szabo, *J. Am. Chem. Soc.* **1982**, *104*, 4546.
[20] G. Lipari, A. Szabo, *J. Am. Chem. Soc.* **1982**, *104*, 4559.
[21] A. Horská, J. Horsky, R. G. S. Spencer, *J. Magn. Reson. Ser. A* **1994**, *110*, 82.
[22] J. Liu, Y. Gong, O. Prakash, L. Wen, I. Lee, J. K. Huang, R. Krishnamoorthi, *Protein Sci.* **1998**, *7*, 132.
[23] A. Heitz, O. Avrutina, D. Le-Nguyen, U. Diederichsen, J. F. Hernandez, J. Gracy, H. Kolmar, L. Chiche, *BMC Struct. Biol.* **2008**, *8*, 54.
[24] J. Liu, O. Prakash, M. Cai, Y. Gong, Y. Huang, L. Wen, J. J. Wen, J. K. Huang, R. Krishnamoorthi, *Biochemistry* **1996**, *35*, 1516.
[25] B. Szenthe, Z. Gaspari, A. Nagy, A. Perczel, L. Graf, *Biochemistry* **2004**, *43*, 3376.
[26] S. M. Pascal, T. Yamazaki, A. U. Singer, L. E. Kay, J. D. Forman-Kay, *Biochemistry* **1995**, *34*, 11353.
[27] A. G. Palmer III, *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 129.
[28] O. Avrutina, H. U. Schmoldt, D. Gabrijelcic-Geiger, D. Le Nguyen, C. P. Sommerhoff, U. Diederichsen, H. Kolmar, *Biol. Chem.* **2005**, *386*, 1301.
[29] J. Austin, W. Wang, S. Puttamadappa, A. Shekhtman, J. A. Camarero, *ChemBioChem* **2009**, *10*, 2663.
[30] K. Dutta, H. Shi, E. R. Cruz-Chu, K. Kami, R. Ghose, *Biochemistry* **2004**, *43*, 8094.
[31] R. Ghose, A. Shekhtman, M. J. Goger, H. Ji, D. Cowburn, *Nat. Struct. Biol.* **2001**, *8*, 998.
[32] S. Arumugam, G. Gao, B. L. Patton, V. Semenchenko, K. Brew, S. R. Van Doren, *J. Mol. Biol.* **2003**, *327*, 719.
[33] N. Trbovic, J. H. Cho, R. Abel, R. A. Friesner, M. Rance, A. G. Palmer III, *J. Am. Chem. Soc.* **2009**, *131*, 615.

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

Backbone Dynamics of Cyclotide MCoTI-I Free and Complexed with Trypsin**

*Shadakshara S. Puttamadappa, Krishnappa Jagadish, Alexander Shekhtman, and Julio A. Camarero**

anie_201002906_sm_miscellaneous_information.pdf

Supplemental Information

General materials and methods

Analytical HPLC was performed on a HP1100 series instrument with 220 nm and 280 nm detection using a Vydac C18 column (5 µm, 4.6 x 150 mm) at a flow rate of 1 mL/min. Semipreparative HPLC was performed on a Waters Delta Prep system fitted with a Waters 2487 Ultraviolet-Visible (UV-vis) detector using a Vydac C18 column (15-20 µm, 10 x 250 mm) at a flow rate of 5 mL/min. All runs used linear gradients of 0.1% aqueous trifluoroacetic acid (TFA, solvent A) vs. 0.1% TFA, 90% acetonitrile in H_2O (solvent B). UV-vis spectroscopy was carried out on an Agilent 8453 diode array spectrophotometer. Electro-spray mass spectrometry was performed on an Applied Biosystems API 3000 triple quadrupole mass spectrometer. Calculated masses were obtained by using ProMac v1.5.3. Protein samples were analyzed by SDS-PAGE on Invitrogen (Carlsbad, CA) 4-20% Tris-Glycine Gels. The gels were then stained with Pierce (Rockford, IL) Gelcode Blue, photographed/digitized using a Kodak (Rochester, NY) EDAS 290, and quantified using NIH Image-J software (http://rsb.info.nih.gov/ij/). DNA sequencing was performed by DNA Sequencing and Genetic Analysis Core Facility at the University of Southern California using an ABI 3730 DNA sequencer, and the sequence data was analyzed with DNAStar (Madison, WI) Lasergene v8.0.2. All chemicals were obtained from Sigma-Aldrich (Milwaukee, WI) unless otherwise indicated.

Construction of MCoTI-I expression plasmid

Plasmids expressing the MCoTI-I linear precursor was constructed using the pTXB1 expression plasmids (New England Biolabs), which contain an engineered *Mxe* Gyrase intein, respectively, and a chitin-binding domain (CBD) as previously described [1].

Expression and purification of recombinant proteins

MCoTI-I cyclotide were produced and characterized as previously described [1]. Briefly, BL21(DE3) cells (Novagen, San Diego, CA) were transformed with the MCoTI-I plasmid (see above). Expression was carried out in M9 minimal medium (6 x 1 L) containing 0.1% $^{15}NH_4Cl$ and 100 mg/ml ampicillin as previously described [1]. Briefly, 5 mL of an overnight starter culture derived from a single clone was used to inoculate 1 L of M9 media. Cells were grown to an OD at 600 nm of ≈ 0.5 at 37°C, and expression was induced by adding isopropyl-β-D-thio-galactopyranoside (IPTG) to a final concentration of 0.3 mM at 30° C for 4 h. The cells were then harvested by centrifugation. For fusion protein purification, the cells were resuspended in 30 mL of lysis buffer (0.1 mM EDTA, 1 mM PMSF, 50 mM sodium phosphate, 250 mM NaCl buffer at pH 7.2 containing 5% glycerol) and lysed by sonication. The lysate was clarified by centrifugation at 15,000 rpm in a Sorval SS-34 rotor for 30 min. The clarified supernatant was incubated with chitin-beads (1 mL beads/L cells, New England Biolabs), previously equilibrated with column buffer (0.1 mM EDTA, 50 mM sodium phosphate, 250 mM NaCl buffer at pH 7.2) at 4° C for 1 h with gentle rocking. The beads were extensively washed with 50 bead-volumes of column buffer containing 0.1% Triton X100 and then rinsed and equilibrated with 50 bead-volumes of column buffer.

Concomitant cleavage, cyclization and folding of [^{15}N]-MCoTI-I cyclotide with reduced glutathione (GSH)

Purified [^{15}N]-MCoTI–Intein-CBD fusion protein was cleaved with 50 mM GSH in degassed column buffer as previously described [1]. The cyclization/folding reactions were kept for up to 2 days at 25°C with gentle rocking. The supernatant of the cyclization reaction was separated by filtration and the beads were washed with additional column buffer (1 column volume per each mL of beads). The supernatant and washes were pooled, and the oxidized-cyclotide was purified by semipreparative HPLC using a linear gradient of 15-45% solvent B over 30 min

yielding ≈1.8 mg of [^{15}N]-MCoTI-I (≈ 0.3 mg/L). Purified product was characterized by HPLC and ES-MS (Fig. S1).

Purification of MCoTI-I using trypsin-sepharose beads.

Preparation of trypsin-sepharose beads: NHS-activated Sepharose was washed with 15 volumes of ice-cold 1 mM HCl. Each volume of beads was incubated with an equal volume of coupling buffer (50 mM NaCl, 200 mM sodium phosphate buffer at pH 6.0) containing 2 mg of porcine pancreas trypsin type IX-S (14,000 units/mg) for 3 h with gentle rocking at room temperature. The beads were then rinsed with 10 volumes of coupling buffer, and incubated with excess coupling buffer containing 100 mM ethanolamine (Eastman Kodak) for 3 h with gentle rocking at room temperature. Finally, the beads were washed with 50 volumes of wash buffer (200 mM sodium acetate buffer at pH 3, 250 mM NaCl) and stored in one volume of wash buffer. ≈30 mL of clarified lysates (in vivo obtained MCoTI-I) or 10 mL of GSH-induced cyclization/folding reaction mixture (in vitro obtained MCoTI-I) were typically incubated with 1.0 - 2.0 mL of trypsin-sepharose for one hour at room temperature with gentle rocking, and centrifuged at 3000 rpm for 1 min. The beads were washed with 50 volumes of PBS containing 0.1% Triton X100, then rinsed with 50 volumes of PBS, and drained of excess PBS. Bound peptides were eluted with 2.0 mL of 8 M GdmHCl at pH ≈4.0 and fractions were desalted and analyzed by RP-HPLC, ES-MS/MS and NMR (Fig. S1).

NMR Spectroscopy

NMR samples of free MCoTI-I were prepared by dissolving [^{15}N]-MCoTI-I into 10 mM potassium phosphate buffer in 90% H_2O/10% ^2H_2O (v/v) at pH ≈ 7.0 to a concentration of 0.2 mM. The trypsin-MCoTI-I complex sample was prepared by titrating trypsin (porcine pancreas trypsin type IX-S, 14,000 units/mg) into 0.2 mM [^{15}N]-MCoTI-I solution. Complex formation was monitored by NMR spectroscopy. No changes in the NMR spectrum were seen after the molar ratio between trypsin and MCoTI-I reached 1:1. The protease inhibitor 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF) was added to inhibit the residual protease

activity of unbound trypsin. Long term stability of the NMR sample was monitored by the NMR spectra of trypsin-[^{15}N]-MCoTI-I. No changes in the NMR peak positions of trypsin-[^{15}N]-MCoTI-I were observed over the period of two months. We used competition experiment of trypsin-[^{15}N]-MCoTI-I with unlabeled MCoTI-I to exclude possibility that trypsin bound cyclotide was cleaved or had scrambled disulfide bonds (Figure S3). By adding ten times molar excess of unlabeled MCoTI-I into trypsin-[^{15}N]-MCoTI-I NMR sample we were able to reconstitute the NMR spectrum of free [^{15}N]-MCoTI-I (Figure S3C). This experiment proved that [^{15}N]-MCoTI-I was not modified by complexing with trypsin.

NMR data were acquired on Bruker Avance II 700 MHz and Avance III 500 MHz spectrometers equipped with ultra-sensitive triple resonance cryoprobes capable of applying pulsed field gradients along the z-axis. All experiments were conducted at 25° C.

Assignments for the amide nitrogens and protons (Table S3) of free and trypsin bound MCoTI-I were obtained by using standard procedures as previously described [1]. Briefly, heteronuclear 3D NMR experiments, ^{1}H{^{15}N}-TOCSY-HSQC and ^{1}H{^{15}N}-NOESY-HSQC, were performed according to standard procedures [8, 9] with spectral widths of 12 ppm in proton dimensions and 35 ppm in nitrogen dimension. The carrier frequency was centered on the water signal, and the solvent was suppressed by using WATERGATE pulse sequence. TOCSY (spin lock time of 80 and 60 ms for free and trypsin bound MCoTI-I, respectively) and NOESY (mixing time of 150 and 80 ms for free and trypsin bound MCoTI-I, respectively) spectra were collected using 1024 t_3 points, 256 t_2 and 128 t_1 blocks of 16 transients. Spectra were processed by using Topspin 1.3 (Bruker). Each 3D-data set was apodized by 90°-shifted sinebell-squared in all dimensions, and zero filled to 1024 x 512 x 256 points prior to Fourier transformation. Spectra were analyzed by using the NMR software program CARA [3].

The measurements of the spin-lattice (R_1) and spin-spin (R_2) relaxation rates, as well as steady state ^{15}N{^{1}H}-nuclear Overhauser effect (NOE) measurements were performed at both 500 MHz and 700 MHz for MCoTI-I in the free and trypsin bound states. The pulse sequence used to

record ^{15}N R_1, R_2 and steady state $^{15}N\{^1H\}$-NOE spectra were used as described [2] with a slight modification to include Watergate techniques for eliminating the water resonance. Decoupling of ^{15}N spins during acquisition was performed using a WALTZ-16 composite pulse sequence with a filed strength of 1.35 kHz. The observed ^1H chemical shifts were determined relative to the internal reference, sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). A recycle delay of 1.5 s was used in the R_1 and R_2 relaxation measurements for both free and trypsin bound samples. The following delays were used to measure the R_2 values at 500 MHz for free MCoTI-I: 20, 30, 50, 60, 80, 100, 140, 300, and 500 ms; at 500 MHz for trypsin bound MCoTI-I: 10, 20, 30, 40, 60, 80, 100, and 300 ms; at 700 MHz for free MCoTI-I: 20, 30, 50, 80, 100, 150, 300, and 5000 ms; and at 700MHz for trypsin bound MCoTI-I: 10, 20, 50, 100, 150, and 300 ms. For R_1 measurements, the following variable relaxation delays were used at 500 MHz, for free MCoTI-I: 20, 60, 100, 200, 400, 500, 700, 1000, and 2000 ms; at 500 MHz for trypsin bound MCoTI-I: 20, 370, 570, 770, 1000, and 1500 ms; at 700 MHz for free MCoTI-I: 20, 50, 200, 370, 520, and 850 ms; and at 700 MHz for trypsin bound MCoTI-I: 20, 50, 100, 250, 500, 770, and 1000 ms. R_2 measurements utilized a 900 μs delay between sequential ^{15}N pulses in the CPMG pulse train for attenuating the ^{15}N signal loss during a R_2 relaxation period. The field strength of the refocusing pulses in the CPMG pulse sequence was 1 kHz.

Heteronuclear steady-state $^{15}N\{^1H\}$-NOE were determined from spectra recorded with (NOE) and without saturation of protons, where saturation was achieved by a train of 120° pulses separated for 5 ms for 1 s. The recycle time between the experiments was 5 s. The NOE measurements were performed using a total of 256 transients per increment in the indirect ^{15}N dimension. The ^{15}N dimension was zero filled to 256 real data points. Heteronuclear NOEs were calculated according to the equation: $\eta = I_{sat}/I_{unsat}$, in which I_{sat} and I_{unsat} are the experimental peak intensities measured from spectra recorded with and without proton saturation. All data were analyzed by using CARA [3]. The R_1 and R_2 relaxation rates were determined by fitting the peak intensity $I(t)$ to a single–exponential function given by $I(t)=I_0e^{-Rt}$,

where *t* is the NMR time delay, using MATLAB program RELAXFIT [4]. Hydrodynamic parameters of free and trypsin bound MCoTI-I were calculated by using RotDif [7]. Residues possessing low NOE values, less than 0.6 for free MCoTI-I and less than 0.2 for trypsin bound MCoTI-I, or participating in slow μs-ms exchanges were removed from calculations. The anisotropy of rotational diffusion for free and trypsin bound MCoTI-I was 2.8 ± 0.3 and 1.5 ± 0.2, respectively [7]. The overall correlation times, τ_c, for free and trypsin bound MCoTI-I was 2.4 ± 0.2 ns and 12.52 ± 0.5 ns, respectively [7], which are consistent with the increased molecular weight of trypsin bound MCoTI-I. Analysis of the micro-dynamic motional parameters using the Lipari-Szabo formalism was performed utilizing DYNAMICS [5, 6] and R_1, R_2 and $^{15}N\{^1H\}$-NOE data at 500 MHz and 700 MHz. Errors (68.3% confidence limits) in the micro-dynamic parameters were obtained from the analytic inverse covariance matrixes of the fits [4]. This method of error estimation is crucial because it takes into account both the random errors as well as the model selection errors.

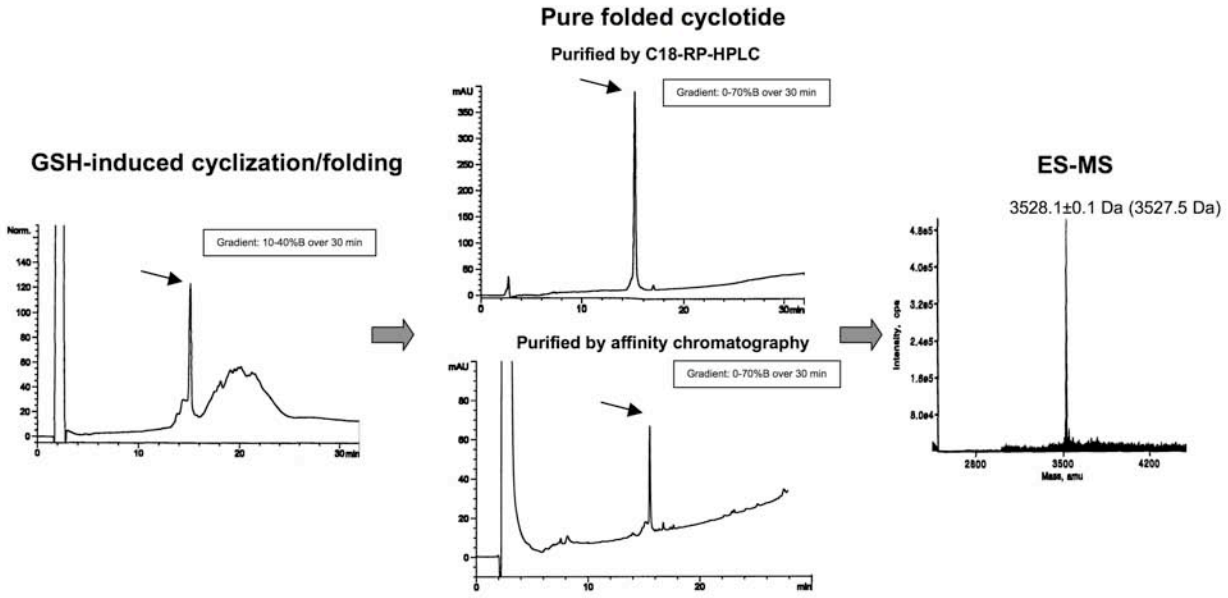


Figure S1. Analytical reversed-phase HPLC trace of GSH-induced cyclization of [^{15}N]-MCoTI-I precursor before and after being purified by preparative HPLC and by affinity chromatography using trypsin-agarose beads. Identification of the folded cyclotide was carried out by ES-MS. Expected molecular weight for the [^{15}N]-MCoTI-I is shown in parenthesis. Natively folded MCoTI-I is indicated with an arrow.

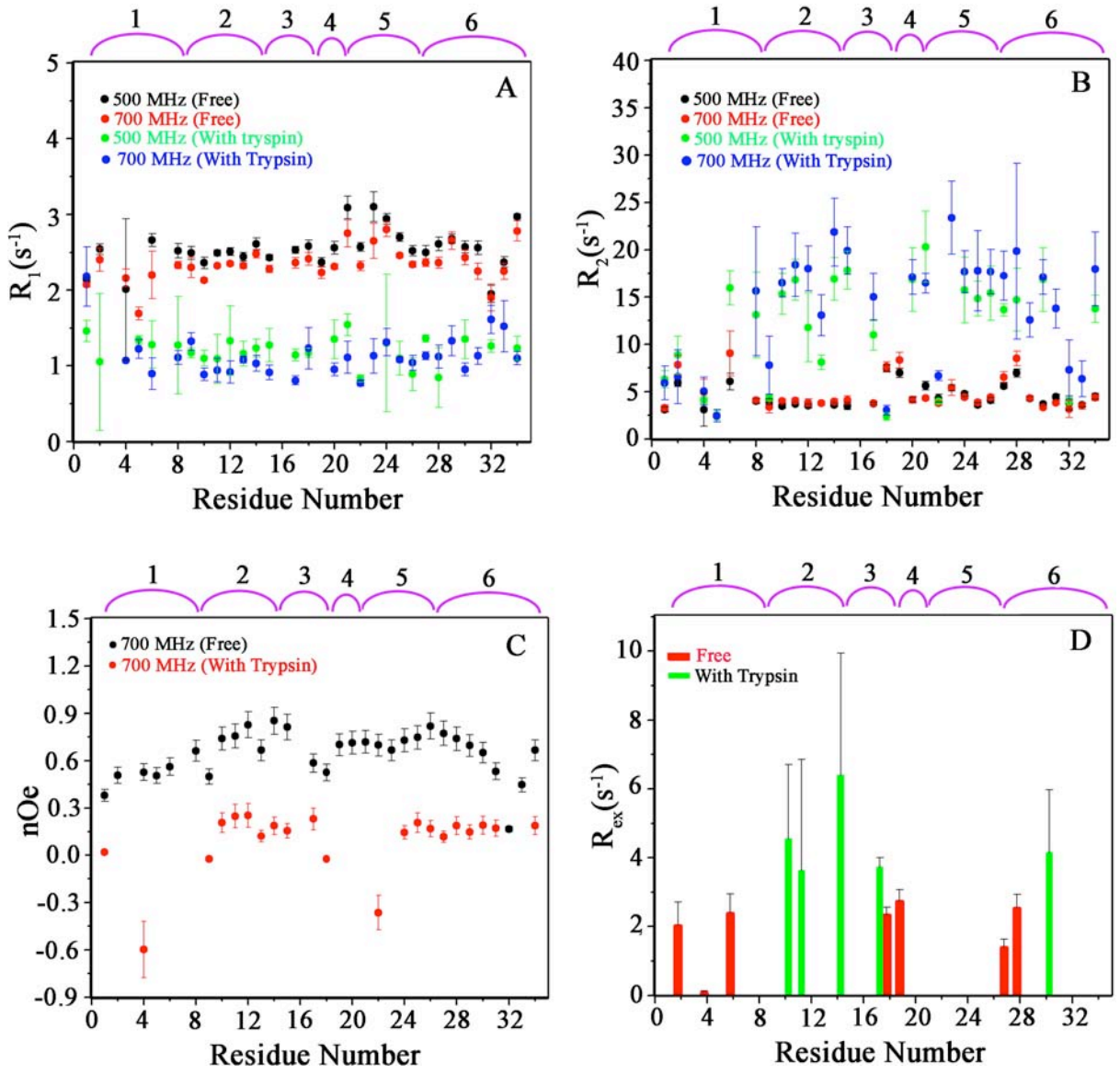


Figure S2. Backbone dynamics of free and trypsin bound MCoTI-I. (A) Spin-lattice relaxation rate, R_1 (s^{-1}), of the free MCoTI-I at 500 MHz (black) at 700 MHz (red) and the trypsin bound MCoTI-I at 500 MHz (green) and 700 MHz (blue). (B) Spin-spin relaxation rate, R_2 (s^{-1}), of the free MCoTI-I at 500 MHz (black) and 700 MHz (red) and the trypsin bound MCoTI-I at 500 MHz (green) and 700 MHz (blue). (C) Heteronuclear nuclear Overhauser effect, $^{15}N\{^1H\}$-NOE, of the free MCoTI-I (black) and the trypsin bound MCoTI-I at 700 MHz. (D) R_{ex} values of the free (red) and trypsin bound MCoTI-I from the Lipari-Szabo analysis are shown. Large R_{ex} values (> 0.1 s^{-1}) suggests that the residue undergoes slow ms-µs motion. The position of MCoTI-I loops are indicated by arabic numbers.

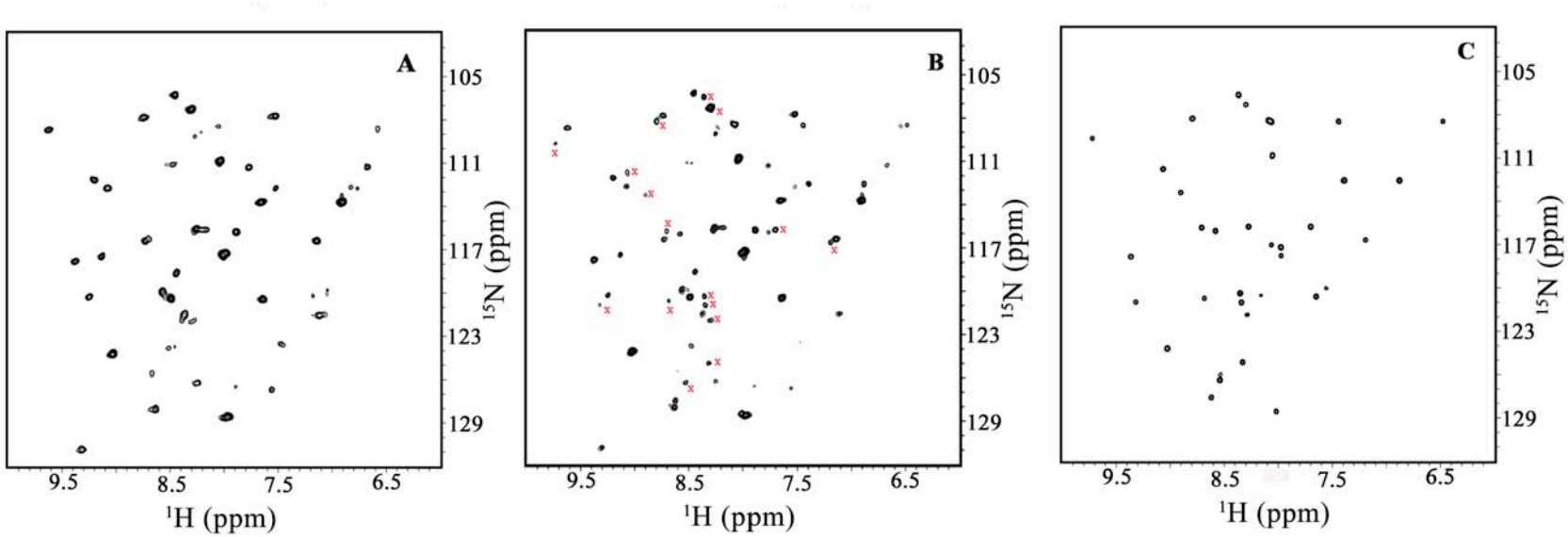


Figure S3 Competition experiment between trypsin-[^{15}N]-MCoTI-I complex and unlabelled MCoTI-I. (A) {^{15}N,^{1}H}-HSQC spectrum of 50 µM trypsin-^{15}N MCoTI-I complex with the molar ratio of 1.1:1. (B) {^{15}N,^{1}H}-HSQC spectrum of 50 µM trypsin-^{15}N-MCoTI-I after adding 80 µM of unlabelled MCoTI-I. The spectrum shows the peaks corresponding to both both free and trypsin bound [^{15}-N]-MCoTI-I. Peaks corresponding to free ^{15}N MCoTI-I are labeled by red crosses. (C) {^{15}N,^{1}H}-HSQC spectrum of 50 µM of trypsin-[^{15}-N]-MCoTI-I after adding 500 µM of unlabelled MCoTI-I. The resultant spectrum is identical to the free ^{15}N-MCoTI-I spectrum, suggesting that [^{15}N]-MCoTI-I was not cleaved by trypsin and MCoTI-I disulfide bonds were intact. NMR spectra were collected at 25^{o}C on 700 MHz Avanve II (Bruker) spectrometer.

Table S1. Free and trypsin bound MCoTI-I order parameters and spectral density function (SDF) models.

Residue #	Free MCoTI-I S^{2*}	Free MCoTI-I SDF-Model	Free MCoTI-I χ^2	Trypsin bound MCoTI-I S^{2*}	Trypsin bound MCoTI-I SDF-Model	Trypsin bound MCoTI-I χ^2	ΔS^2
1, Val	0.68	LS[a]	24.68	0.34	CL[c]	6.23	0.35
2, Cys	0.79	LS-ex[b]	4.92				
4, Lys	0.85	LS-ex	6.46	0.23	CL	3.72	0.61
6, Leu	0.70	CL-ex[d]	0.26	0.75	LS	12.45	-0.05
8, Arg	0.88	LS	6.11				
9, Cys	0.80	LS	3.53	0.23	CL	7.63	0.57
10, Arg	0.81	LS	6.33	0.51	CL-ex	3.81	0.31
11, Arg	0.87	LS	25.49	0.74	LS	5.27	0.13
12, Asp	0.89	LS	27.85	0.79	LS	2.38	0.10
13, Ser	0.85	LS	23.67	0.43	CL	3.62	0.43
14, Asp	0.89	LS	20.76	0.62	CL-ex	2.44	0.27
15, Cys	0.82	LS	21.65	0.81	LS	1.44	0.01
17, Gly	0.83	LS	6.98	0.48	CL-ex	0.03	0.35
18, Ala	0.85	LS	17.91				
19, Cys	0.79	LS	8.24				
20, Ile	0.88	LS	3.54	0.77	LS	3.12	0.11
21, Cys	0.98	LS	13.31	0.54	CL	2.53	0.44
22, Arg	0.86	LS	5.12	0.18	CL	16.40	0.68
24, Asn	0.99	LS	10.22	0.73	LS	9.00	0.26
25, Gly	0.92	LS	28.68	0.75	LS	32.64	0.17
26, Tyr	0.93	CL	7.97	0.78	LS	10.90	0.15
27, Cys	0.88	LS-ex	16.51	0.79	LS	29.22	0.09
28, Gly	0.89	LS-ex	3.49	0.75	LS	5.74	0.14
29, Ser	0.91	LS	9.45				
30, Gly	0.79	LS	39.23	0.58	CL-ex	1.68	0.21
31, Ser	0.81	LS	7.09				
32, Asp	0.56	LS	13.41				
33, Gly	0.51	LS	6.85				
34, Gly	0.98	LS	4.99	0.77	LS	4.20	0.21

[a] LS corresponds to the original Lipari-Szabo SDF model [5, 10], assuming that the characteristic correlation time for local fluctuations, $\tau_{loc,}$ is much shorter than the global rotational correlation time, τ_c, [10].
[b] LS-ex corresponds to the Lipari-Szabo SDF model supplemented with the conformational exchange term [10].
[c] CL corresponds to the extended Lipari-Szabo SDF model [11].
[d] CL-ex corresponds to the extended Lipari-Szabo SDF model supplemented with conformational exchange term [11].
* The order parameter S^2 was calculated by minimizing $\chi^{2.}$ [5], which includes relaxation data set collected at 500 MHz and 700 MHz frequencies (R_1, R_2 and NOE).

$$\chi^2 = \sum_i \sum_f \left\{ \left(\frac{R_{1i}^{obs} - R_{1i}^{cal}}{\sigma_{R_{1i,f}}} \right)^2 + \left(\frac{R_{2i;f}^{obs} - R_{2i;f}^{cal}}{\sigma_{R_{2i,f}}} \right)^2 + \left(\frac{NOE_{i;f}^{obs} - NOE_{i;f}^{cal}}{\sigma_{NOE_{i,f}}} \right)^2 \right\}$$

Where the sum runs over all the residues as well as over the frequencies (f = 500 MHz and 700 MHz)
$\sigma_{R1i,f}$, $\sigma_{R2i,f}$ and $\sigma_{NOEi,f}$ are standard deviations in $R_{1i,f}$, $R_{2i,f}$ and $NOE_{i,f}$ for the *i*th residue, respectively,
The superscripts refer to the observed (obs) and calculated (cal) [10,11] values of the relaxation parameters.

Table S2. Summary of the ^{1}H and $^{15}N^{\alpha}$-^{1}H NMR assignments for the backbone protons (i.e. N^{α}-H) of free and trypsin bound MCoTI-I.

Residue #	Free δ_{NH}(ppm)	Trypsin bound δ_{NH}(ppm)	$\Delta\delta_{NH}$(ppm)	Free δ_{N}(ppm)	Trypsin bound δ_{N}(ppm)	$\Delta\delta_{N}$(ppm)	$[(\Delta\delta)^2_{NH}+0.04\Delta\delta^2_{N})/2]^{1/2}$
1, Val	8.34	8.57	-0.23	120.97	124.80	-3.83	0.57
2, Cys	8.54	8.09	0.45	126.33	126.41	-0.08	0.32
4, Lys	8.08	8.45	-0.37	120.51	123.74	-3.23	0.53
5, Ile	7.52	7.86	-0.34	119.94	118.51	1.43	0.31
6, Leu	8.52	8.24	0.28	125.77	126.23	-0.46	0.21
8, Arg	8.62	9.31	-0.69	127.56	127.9	-0.34	0.49
9, Cys	8.35	8.38	-0.03	120.36	121.92	-1.56	0.22
10, Arg	7.97	7.98	-0.01	117.13	117.27	-0.14	0.02
11, Arg	9.36	9.37	-0.01	117.84	117.83	0.01	0.01
12, Asp	9.32	9.24	0.08	120.97	120.31	0.66	0.11
13, Ser	8.28	8.25	0.03	115.76	115.57	0.19	0.03
14, Asp	7.65	7.63	0.02	120.59	120.43	0.16	0.03
15, Cys	7.97	8.43	-0.46	117.74	118.64	-0.90	0.35
17, Gly	8.37	8.45	-0.08	106.61	106.26	0.35	0.08
18, Ala	8.33	7.88	0.45	125.08	126.84	-1.76	0.41
19, Cys	8.06	8.04	0.02	117	116.20	0.80	0.11
20, Ile	8.90	9.09	0.19	113.3	112.8	-0.5	0.01
21, Cys	9.02	9.0	0.00	124.19	124.27	-0.08	0.01
22, Arg	8.02	7.95	0.07	128.56	128.54	0.02	0.04
23, Gly	8.79	8.73	0.06	108.26	107.79	0.47	0.08
24, Asn	7.70	7.88	-0.18	115.74	115.76	-0.02	0.13
25, Gly	8.30	8.29	0.01	107.30	107.23	0.07	0.01
26, Tyr	7.19	7.14	0.05	116.62	116.39	0.23	0.05
27, Cys	8.68	8.48	0.20	120.71	120.41	0.30	0.15
28, Gly	9.72	9.61	0.11	109.64	108.66	0.98	0.16
29, Ser	8.71	8.69	0.02	115.75	116.10	-0.35	0.05
30, Gly	9.06	9.06	0.00	111.77	112.71	-0.94	0.13
31, Ser	8.57	8.72	-0.15	116.00	116.40	-0.40	0.12
32, Asp	8.31	8.31	0.00	122.06	122.05	0.01	0.01
33, Gly	8.08	8.04	0.04	108.41	108.42	-0.01	0.02
34, Gly	8.05	8.02	0.03	110.78	110.84	-0.06	0.02

References

[1] a) J. A. Camarero, R. H. Kimura, Y. H. Woo, A. Shekhtman, J. Cantor, Chembiochem. **2007** Aug 13;8(12):1363-6; b) J. Austin, W. Wang, S. Puttamadappa, A. Shekhtman, J. A. Camarero, *Chembiochem* **2009**, *10*, 2663.

[2] N. A. Farrow, R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay, L. E. Kay, *Biochemistry* **1994**, *33*, 5984.

[3] J. E. Masse, R. Keller, K. Pervushin, *J. Magn. Reson.* **2006**, *181*, 45.

[4] J. Blake-Hall, O. Walker, D. Fushman, *Methods Mol. Biol.* **2004**, *278*, 139.

[5] D. Fushman, S. Cahill, D. Cowburn, *J. Mol. Biol.* **1997**, *266*, 173.

[6] G. Lipari, A. Szabo, *Biochemistry* **1981**, *20*, 6250.

[7] O. Walker, R. Varadan, D. Fushman, *J. Magn. Reson.* **2004**, 168, 336.

[8] J. Cavanagh, M. Rance, *J. Magn. Res.* **1992**, *96*, 670.

[9] K. Wuthrich, *NMR of Proteins and Nucleic Acids.*, **1986.**

[10] G. Lipari, A. Szabo, *J. Am. Chem. Soc.* **1982**, *104*, 4559.

[11] G. Clore, A. Szabo, A. Bax, L.E. Kay, P. Driscoll, P. Wingfield, A. Gronenborn, *J. Am. Chem. Soc.* **1990**, *112*, 4989.

Review
Cyclotides, A Promising Molecular Scaffold for Peptide-Based Therapeutics

Krishnappa Jagadish, Julio A. Camarero
Department of Pharmacology and Pharmaceutical Sciences, University of Southern California, Los Angeles, CA 90033

Received 21 December 2009; revised 23 February 2010; accepted 7 March 2010
Published online 26 May 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.21433

ABSTRACT:

Cyclotides are a new emerging family of large plant-derived backbone-cyclized polypeptides (≈30 amino acids long) that share a disulfide-stabilized core (three disulfide bonds) characterized by an unusual knotted structure. Their unique circular backbone topology and knotted arrangement of three disulfide bonds make them exceptionally stable to thermal, chemical, and enzymatic degradation compared to other peptides of similar size. Currently, more than 100 sequences of different cyclotides have been characterized, and the number is expected to increase dramatically in the coming years. Considering their stability and biological activities like anti-HIV, uterotonic, and insecticidal, and also their abilities to cross the cell membrane, cyclotides can be exploited to develop new stable peptide-based drugs. We have recently demonstrated the intriguing possibility of producing libraries of cyclotides inside living bacterial cells. This opens the possibility to generate large genetically encoded libraries of cyclotides that can then be screened inside the cell for selecting particular biological activities in a high-throughput fashion. The present minireview reports the efforts carried out toward the selection of cyclotide-based compounds with specific biological activities for drug design. © 2010 Wiley Periodicals, Inc. Biopolymers (Pept Sci) 94: 611–616, 2010.

Keywords: *cyclotide; libraries; cell-based screening*

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version. You can request a copy of the preprint by emailing the Biopolymers editorial office at biopolymers@wiley.com

INTRODUCTION

Peptides have proven to be valuable and effective drugs when targeting protein–protein interactions.[1,2] The concept of using peptides to modulate intracellular processes has been investigated for decades, as peptides play a central role in every cell in the body. Polypeptides that mimic protein fragments typically are able to provide effective competitive-binding antagonists for protein–protein interactions. Targeting these interactions, which, usually involve large binding surfaces, has proven challenging for small molecules. The utility of peptide therapeutics, however, has typically been limited by their generally poor stability and limited bioavailability. For example, linear peptides are typically inherently unstable within the body and are rapidly broken down into inactive fragments by proteolytic enzymes, which are then filtered from the blood stream by the kidneys within minutes. In response to this challenge, a number of novel polypeptide scaffolds are starting to emerge to replace classical peptide and protein-based therapeutics.[2–9]

Correspondence to: Julio A. Camarero, Associate Professor, Department of Pharmacology and Pharmaceutical Sciences, University of Southern California, 1985 Zonal Avenue, PSC 616, Los Angeles, CA 90033, USA; e-mail: jcamarer@usc.edu
Contract grant sponsor: School of Pharmacy at the University of Southern California
Contract grant sponsor: National Institute of Health
Contract grant number: GM090323-01

This article reviews the latest developments on the use of cyclotides as a molecular scaffold for delivering a novel type of peptide-based therapeutics with the specificity of proteins but the physical–chemical properties that are usually associated with small molecule-based drugs.

Cyclotides, a Novel Ultrastable Polypeptide Scaffold

Special attention has been recently given to the use of highly constrained peptides, also known as micro- or miniproteins, as extremely stable and versatile scaffolds for the production of high-affinity ligands for specific protein capture and/or development of therapeutics.[5,8] Cyclotides are fascinating microproteins ($\approx$30 amino acids long) present in plants from the Violaceae, Rubiaceae, and also Cucurbitaceae; and featuring various biological actions such as protease inhibitory, insecticidal, cytotoxic, anti-HIV, or hormonelike activity.[10,11] More recently, it has been demonstrated that cyclotides also show potent anthelmintic activity.[12,13] Their insecticidal and anthelmintic properties suggest that they may function as defense molecules in plants.

Cyclotides share a unique head-to-tail circular knotted topology of three disulfide bridges, with one disulfide penetrating through a macrocycle formed by the two other disulfides and interconnecting peptide backbones, forming what is called a cystine knot topology (see Figure 1). These microproteins can be considered as natural combinatorial peptide libraries structurally constrained by the cystine-knot scaffold and head-to-tail cyclization but in which hypermutation of essentially all residues is permitted with the exception of the strictly conserved cysteines that comprise the knot.[17–19] The main features of cyclotides are therefore a remarkable stability due to the cystine knot, a small size making them readily accessible to chemical synthesis, and an excellent tolerance to sequence variations. For example, the first cyclotide to be discovered, kalata B1 (kB1), is an orally effective uterotonic.[20] Intriguingly, the MCoTI-II cyclotide has also been shown to cross the cell membrane through macropinocytosis.[21]

Using the Cyclotide Scaffold to Introduce Novel Biological Activities

There have been a number of reports showing the plasticity of the cyclotide framework and its tolerance to substitution. The first proof of concept for grafting of new sequences onto

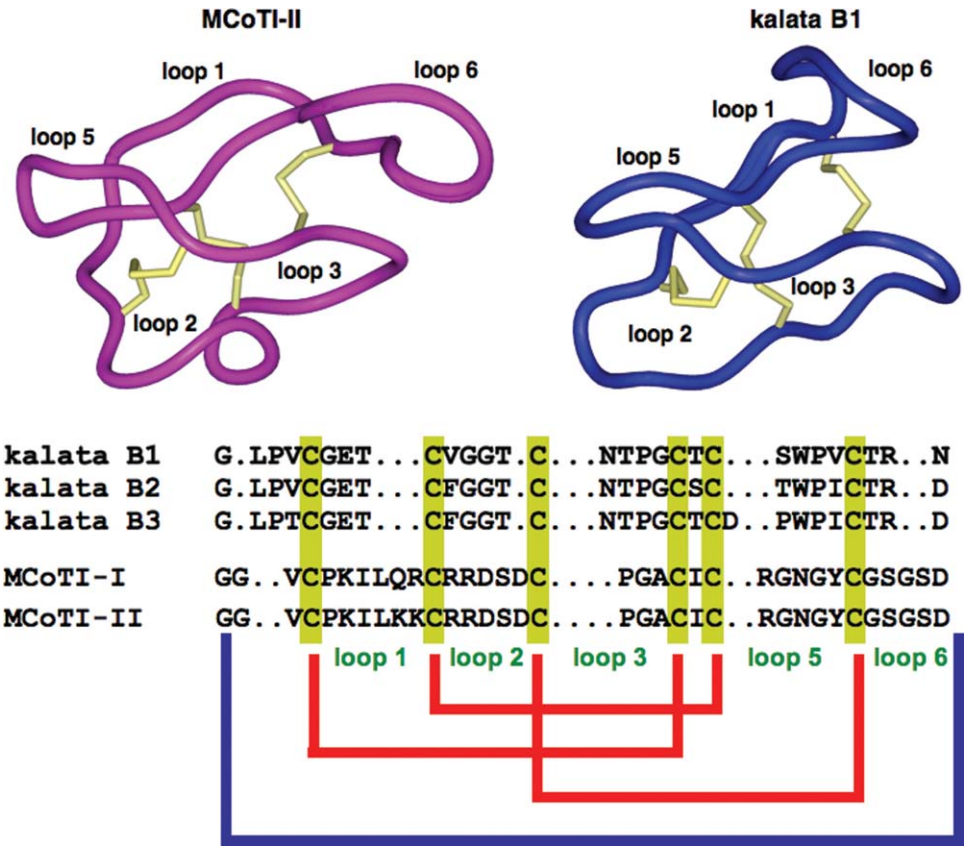


FIGURE 1 Primary and tertiary structure of cyclotides from the plants *Momordica cochinchinensis* (MCoTI-II) and *Oldenlandia affinis* (kalata B1).[14–16] Red and blue connectors indicate backbone and disulfide bonds, respectively.

kB1 was established by replacing some hydrophobic residues in loop 5 with charged and polar residues. These residues form part of a surface-exposed hydrophobic patch that plays a significant role in the folding and biological activity of kB1. The modified cyclotide analogs retained the native fold and lacked the undesirable hemolytic activity of the parent peptide despite the importance of these residues.[17]

Novel VEGF-A antagonists have also been developed by grafting a peptide sequence able to antagonize VEGF-A onto kB1.[22] One of the grafted analogs showed biological activity at low-micromolar concentration in an in vitro VEGF-A antagonism assay, and the in vitro stability of the target epitope was markedly increased using this approach.

The use of the cyclotide scaffold in drug design has been also recently shown by engineering non-native activities into the cyclotide MCoTI-II (see Figure 1).[23] Replacing the P1 residue in the active loop of the cyclotide produced several MCoTI-II analogs with different specificities toward alternative protease targets. Interestingly, several analogs showed selective low-μM inhibition of foot-and-mouth-disease virus 3C protease. These are the first-reported peptide-based inhibitors of this protease and although the potency was relatively low (low μM), this study demonstrates the potential of using MCoTI-based cyclotides for designing novel protease inhibitors.[23] Hence, the MCoTI-cyclotide appears to be a versatile scaffold for the display of biological activities as it has also been shown to be able to cross cellular membranes.[21] This opens the possibility of using it for delivering biological active peptide sequences to intracellular targets. Several grafting studies using the MCoTI-cyclotide scaffold to target several intracellular targets involved in programmed cell death and in tumor cell proliferation and suppression are currently underway in our laboratory.

MCoTI cyclotides share a high-sequence homology with related cystine-knot trypsin inhibitors found in squash such as EETI-II (*Ecballium elaterium* trypsin inhibitor II), and, in fact, can be considered cyclized homologs of the these protease inhibitors. Squash cystine-knot trypsin inhibitors have also been successfully used to graft biological activities. Thus, the RGD sequence, originally discovered in dis-integrins, has been grafted into loop 1 of EETI-II yielding an EETI-II analog with platelet inhibitory activity.[24] Interestingly, the engineered proteins were much more potent in inhibiting platelet aggregation than the grafted peptides, highlighting the importance of grafting a linear epitope into a stable peptide-scaffold. These highly stable peptides could have clinical use for the treatment of patients with acute coronary syndrome, for example.

In addition to displaying biological activities, EETI-II peptides have also been shown to permeate through rat small intestinal mucose more effectively compared to other peptide drugs such as insulin and bacitracin.[25] These analogs also showed comparable stability to cyclotides in plasma, thus supporting evidence for the utility of highly constrained cystine-knot peptides in drug design.

All these data highlight the extraordinary pharmacokinetic properties of cyclotides and cystine-knot peptides in general thus confirming the potential of these polypeptide scaffolds in peptide-based drug discovery.

Biosynthesis of Cyclotides

Cyclotides are ribosomally produced in plants from precursors that comprise between one and three cyclotide domains; however, the mechanism of excision of the cyclotide domains and ligation of the free N- and C-termini to produce the circular peptides has not been completely elucidated yet.[26,27]

Our group has recently developed and successfully used a biomimetic approach for the biosynthesis of folded cyclotides inside bacterial cells by making use of modified protein splicing units in combination with an in-cell intramolecular native chemical ligation reaction (NCL) (see Figure 2).[16,28,29] Intramolecular NCL requires the presence of an N-terminal Cys residue and a C-terminal α-thioester group in the same linear precursor molecule.[30–32] For this purpose, the linear cyclotide pre-

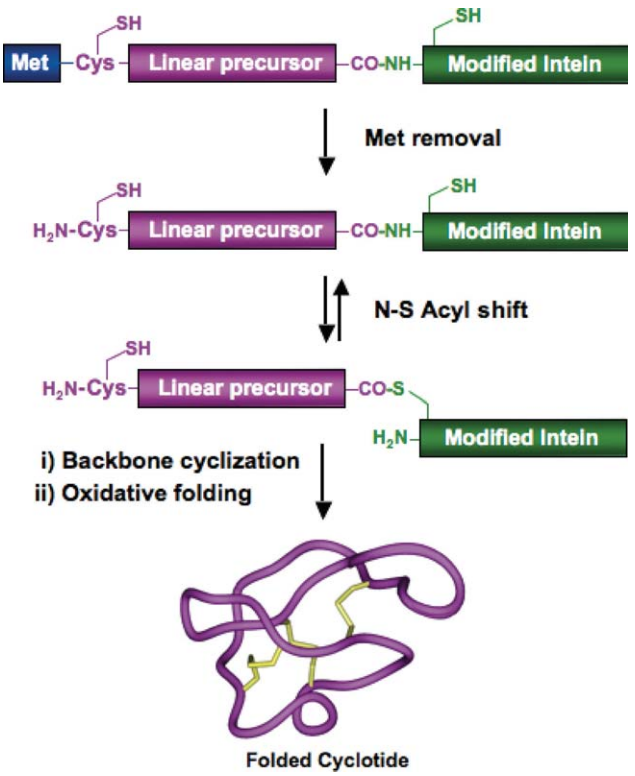


FIGURE 2 Biosynthetic approach for in vivo production of cyclotides kalata B1 and MCoTI-II inside live *E. coli* cells.[18,29] Backbone cyclization of the linear precursor is mediated by a modified protein splicing unit or intein. The cyclized product then folds spontaneously in the bacterial cytoplasm.

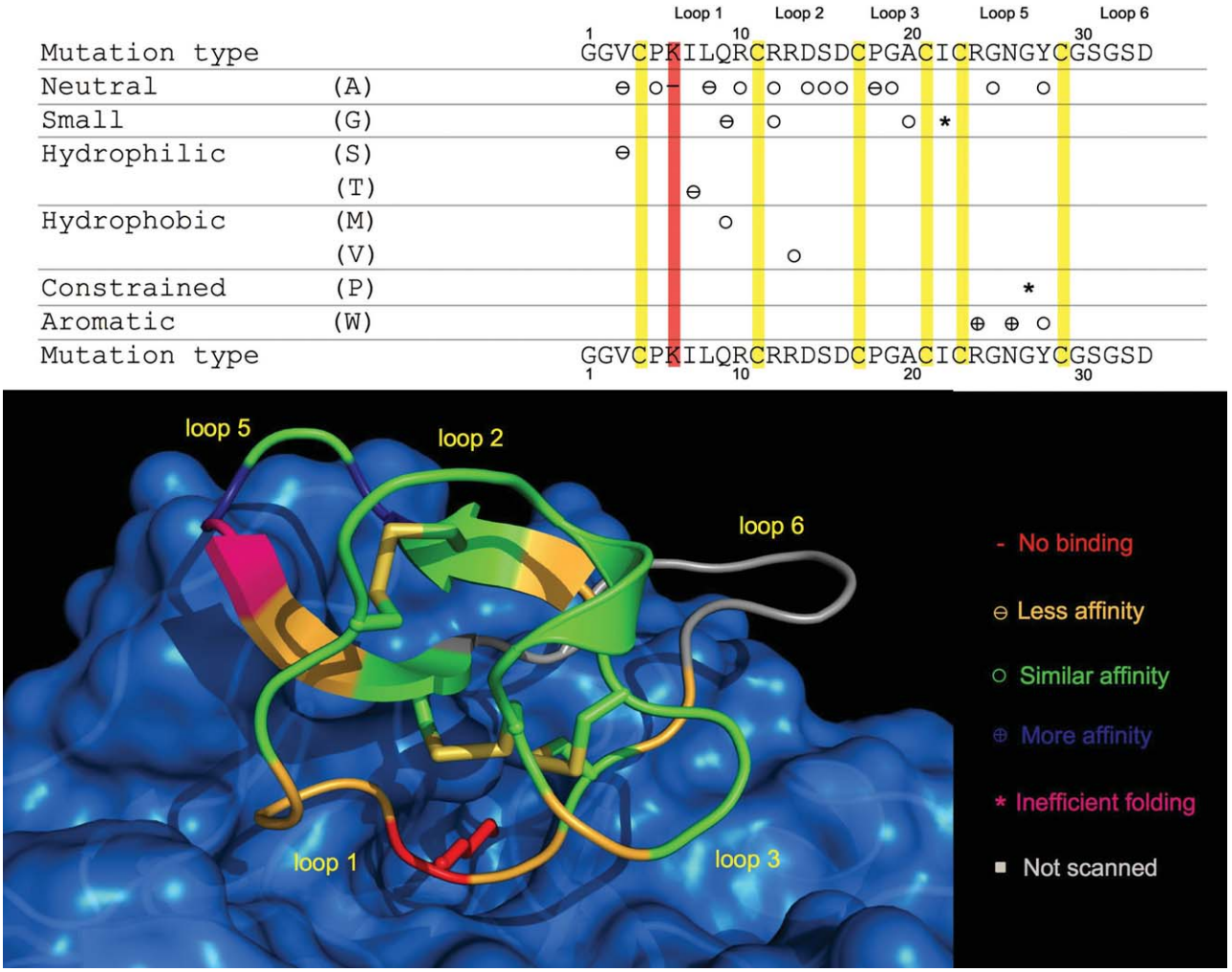


FIGURE 3 Relative affinities for trypsin of a series of MCoTI-I mutants covering all the loop positions except loop 6 and Cys residues. A model of cyclotide MCoTI-I bound to trypsin is shown at the bottom indicating the position of the mutations. The side-chain of residue K6 is shown in red bound to specificity pocket of trypsin. The model was produced by homology modeling at the Swiss model workspace[35] using the structure of CPTI-II-trypsin complex (PDB code: 2btc)[36] as template. Structure was generated using the PyMol software package. Figure adapted from reference 16.

cursors were fused in frame at their C- and N-terminus to a modified intein and a Met residue, respectively. This allows the generation of the required C-terminal α-thioester and N-terminal Cys residue after in vivo processing by endogenous Met amino peptidase. Our group has also recently used this biomimetic approach for the biosynthesis of another backbone-cyclized peptide, the Bowman–Birk inhibitor sunflower trypsin inhibitor 1[33]; and biosynthesis of other cyclic peptides such backbone-cyclized α-defensins and naturally occurring θ-defensins is currently underway in our laboratory.

Recombinant biosynthesis of cyclic polypeptides offers many advantages over purely synthetic methods.[32] Using the tools of molecular biology, large combinatorial libraries of cyclic peptides may be generated and screened in vivo. A typical chemical synthesis may generate 10^4 different molecules. It is not uncommon for a recombinant library to contain as many as 10^9 members. The molecular diversity generated by this approach is analogous to phage-display technology. The approach, however, differs from phage-display in that the backbone-cyclized polypeptides are not fused to or displayed by any viral particle or protein, but remain on the inside of the living cell where they can be further screened for biological activity in an analogous way as the yeast two hybrid technology works.[34] The complex cellular cytoplasm provides the appropriate environment to address the physiological relevance of potential leads.

Screening of Cyclotide-Based Libraries

The ability to create cyclic polypeptides in vivo opens up the possibility of generating large libraries of cyclic polypeptides. Using the tools of molecular biology, genetically encoded libraries of cyclic polypeptides containing billions of members can be readily generated. This tremendous molecular diversity forms the basis for selection strategies that model natural evolutionary processes. Also, because the cyclic polypeptides are generated inside living cells, these libraries can be directly screened for their ability to attenuate or inhibit cellular processes.

We have reported recently the biosynthesis of a genetically encoded library of MCoTI-I based cyclotides in *Escherichia coli* cells.[19] The cyclization/folding of the library was performed either in vitro, by incubation with a redox buffer containing glutathione, or by in vivo self-processing of the corresponding precursor proteins. Of 27 mutations studied, only two mutations, G25P and I20G, negatively affected the folding of cyclotides (Figures 1 and 3). These data provide significant insights into the structural constraints of the MCoTI cyclotide framework and the functional elements for trypsin binding. To our knowledge, this is the first time that the biosynthesis of a genetically encoded library of MCoTI-based cyclotides containing a complete suite of amino acid mutants is reported. Craik and coworkers have also recently reported the chemical synthesis of a complete suite of Ala mutants for kB1.[18] These mutants were fully characterized structurally and functionally. Their results indicated that only two of the mutations explored (kB1 W20A and P21A, both located in loop 5, see Figure 1) prevented folding.[18] The mutagenesis results obtained in our work show similar results highlighting the extreme robustness of the cyclotide scaffold to mutations. These studies show that cyclotides may provide an ideal scaffold for the biosynthesis of large combinatorial libraries inside living bacterial cells, which can then be screened in-cell for biological activity using high-throughput flow cytometry techniques.[37–39]

Our group has recently developed a protease cell-based screening reporter using a couple of optimized genetically encoded fluorescent proteins for FRET-based screening.[39] We are currently using this approach for screening genetically encoded libraries based on the MCoTI-I cyclotide scaffold against several proteases, including Anthrax lethal factor.

CONCLUSION AND REMARKS

In summary, cyclotides have several characteristics that make them ideal drug development tools. First, they are remarkably stable due to the cystine knot. Second, they are small, making them readily accessible to chemical synthesis and therefore chemical lead optimization.[23,40–42] Third, they can be encoded within standard cloning vectors, expressed in bacteria or animal cells, and are amenable to substantial sequence variation.[19,28,29] Finally, some cyclotides have been shown to be orally bioavailable[20] and able to cross the cell membrane through macropinocytosis.[21] These characteristics make them ideal substrates for molecular evolution strategies to enable generation and selection of compounds with optimal binding and inhibitory characteristics. Cyclotides thus appear as very promising leads or frameworks for the development of novel peptide-based therapeutics and diagnostics.[4,5,8,17]

REFERENCES

1. Saladin, P. M.; Zhang, B. D.; Reichert, J. M. ID Rugs 2009, 12, 779–784.
2. Lewis, R. J. Adv Exp Med Biol 2009, 655, 44–48.
3. Craik, D. J.; Simonsen, S.; Daly, N. L. Curr Opin Drug Discov Dev 2002, 5, 251–260.
4. Craik, D. J.; Cemazar, M.; Daly, N. L. Curr Opin Drug Discov Devel 2006, 9, 251–260.
5. Craik, D. J.; Clark, R. J.; Daly, N. L. Expert Opin Investig Drugs 2007, 16, 595–604.
6. Kolmar, H.; Skerra, A. FEBS J 2008, 275, 2667.
7. Skerra, A. FEBS J 2008, 275, 2677–2683.
8. Sancheti, H.; Camarero, J. A. Adv Drug Deliv Rev 2009, 61, 908–917.
9. Bloom, L.; Calabro, V. Drug Discov Today 2009, 14, 949–955.
10. Craik, D. J.; Daly, N. L.; Mulvenna, J.; Plan, M. R.; Trabi, M. Curr Protein Pept Sci 2004, 5, 297–315.
11. Daly, N. L.; Rosengren, K. J.; Craik, D. J. Adv Drug Deliv Rev 2009, 61, 918–930.
12. Colgrave, M. L.; Kotze, A. C.; Ireland, D. C.; Wang, C. K.; Craik, D. J. Chembiochem 2008, 9, 1939–1945.
13. Colgrave, M. L.; Kotze, A. C.; Huang, Y. H.; O'Grady, J.; Simonsen, S. M.; Craik, D. J. Biochemistry 2008, 47, 5581–5589.
14. Saether, O.; Craik, D. J.; Campbell, I. D.; Sletten, K.; Juul, J.; Norman, D. G. Biochemistry 1995, 34, 4147–4158.
15. Hernandez, J. F.; Gagnon, J.; Chiche, L.; Nguyen, T. M.; Andrieu, J. P.; Heitz, A.; Hong, T. T.; Pham, T. T.; Le Nguyen, D. Biochemistry 2000, 39, 5722–5730.
16. Heitz, A.; Hernandez, J. F.; Gagnon, J.; Hong, T. T.; Pham, T. T.; Nguyen, T. M.; Le-Nguyen, D.; Chiche, L. Biochemistry 2001, 40, 7973–7983.
17. Clark, R. J.; Daly, N. L.; Craik, D. J. Biochem J 2006, 394, 85–93.
18. Simonsen, S. M.; Sando, L.; Rosengren, K. J.; Wang, C. K.; Colgrave, M. L.; Daly, N. L.; Craik, D. J. J Biol Chem 2008, 283, 9805–9813.
19. Austin, J.; Wang, W.; Puttamadappa, S.; Shekhtman, A.; Camarero, J. A. Chembiochem 2009, 10, 2663–2670.
20. Gran, L. Lloydia 1973, 36, 174–178.
21. Greenwood, K. P.; Daly, N. L.; Brown, D. L.; Stow, J. L.; Craik, D. J. Int J Biochem Cell Biol 2007, 39, 2252–2264.
22. Gunasekera, S.; Foley, F. M.; Clark, R. J.; Sando, L.; Fabri, L. J.; Craik, D. J.; Daly, N. L. J Med Chem 2008, 51, 7697–7704.

23. Thongyoo, P.; Roque-Rosell, N.; Leatherbarrow, R. J.; Tate, E. W. Org Biomol Chem 2008, 6, 1462–1470.
24. Reiss, S.; Sieber, M.; Oberle, V.; Wentzel, A.; Spangenberg, P.; Claus, R.; Kolmar, H.; Losche, W. Platelets 2006, 17, 153–157.
25. Werle, M.; Schmitz, T.; Huang, H. L.; Wentzel, A.; Kolmar, H.; Bernkop-Schnurch, A. J Drug Target 2006, 14, 137–146.
26. Gillon, A. D.; Saska, I.; Jennings, C. V.; Guarino, R. F.; Craik, D. J.; Anderson, M. A. Plant J 2008, 53, 505–515.
27. Saska, I.; Gillon, A. D.; Hatsugai, N.; Dietzgen, R. G.; Hara-Nishimura, I.; Anderson, M. A.; Craik, D. J. J Biol Chem 2007, 282, 29721–29728.
28. Kimura, R. H.; Tran, A. T.; Camarero, J. A. Angew Chem Int Ed Engl 2006, 45, 973–976.
29. Camarero, J. A.; Kimura, R. H.; Woo, Y. H.; Shekhtman, A.; Cantor, J. Chembiochem 2007, 8, 1363–1366.
30. Camarero, J. A.; Pavel, J.; Muir, T. W. Angew Chem Int Ed 1997, 37, 347–349.
31. Camarero, J. A.; Muir, T. W. J Am Chem Soc 1999, 121, 5597–5598.
32. Camarero, J. A.; Mitchell, A. R. Protein Pept Lett 2005, 12, 723–728.
33. Austin, J.; Kimura, R. H.; Woo, Y. H.; Camarero, J. A. Amino Acids 2010, 38, 1313–1322.
34. Suter, B.; Auerbach, D.; Stagljar, I. Biotechniques 2006, 40, 625–644.
35. Guex, N.; Peitsch, M. C. Electrophoresis 1997, 18, 2714–2723.
36. Helland, R.; Berglund, G. I.; Otlewski, J.; Apostoluk, W.; Andersen, O. A.; Willassen, N. P.; Smalas, A. O. Acta Crystallogr D Biol Crystallogr 1999, 55, 139–148.
37. Giepmans, B. N.; Adams, S. R.; Ellisman, M. H.; Tsien, R. Y. Science 2006, 312, 217–224.
38. You, X.; Nguyen, A. W.; Jabaiah, A.; Sheff, M. A.; Thorn, K. S.; Daugherty, P. S. Proc Natl Acad Sci USA 2006, 103, 18458–18463.
39. Kimura, R. H.; Steenblock, E. R.; Camarero, J. A. Anal Biochem 2007, 369, 60–70.
40. Tam, J. P.; Lu, Y. A.; Yang, J. L.; Chiu, K. W. Proc Natl Acad Sci USA 1999, 96, 8913–8918.
41. Daly, N. L.; Love, S.; Alewood, P. F.; Craik, D. J. Biochemistry 1999, 38, 10606–10614.
42. Thongyoo, P.; Tate, E. W.; Leatherbarrow, R. J. Chem Commun (Camb) 2006, 2848–2850.

DOI: 10.1002/cbic.200900534

Biosynthesis and Biological Screening of a Genetically Encoded Library Based on the Cyclotide MCoTI-I

Jeffrey Austin,[b] Wan Wang,[a] Swamy Puttamadappa,[c] Alexander Shekhtman,[c] and Julio A. Camarero*[a, b]

Introduction

Cyclotides are fascinating microproteins that are present in plants from Violaceae, Rubiaceae and Cucurbitaceae and exhibit various biological properties such as protease inhibitory, antimicrobial, insecticidal, cytotoxic, anti-HIV or hormone-like activity.[1,2] They share a unique head-to-tail circular knotted topology of three disulfide bridges, with one disulfide penetrating through the macrocycle formed by the two other disulfides and interconnecting peptide backbones to form what is called a cystine knot topology (Figure 1 A).

Cyclotides have several characteristics that make them promising leads or frameworks for peptide drug design.[3,4] The cystine knot and cyclic backbone topology makes them exceptionally resistant to thermal, chemical, and enzymatic degradation compared with other peptides of similar size.[5] Some cyclotides have been shown to be orally bioavailable. For example, the first cyclotide to be discovered, Kalata B1, was found to be an orally effective uterotonic,[6] and other cyclotides have been shown to cross the cell membrane through macropinocytosis.[7] Moreover, immunogenicity is generally considered not to be a major issue for small-sized and stable microproteins.[8,9] Cyclotides are also amenable to substantial sequence variation and they can be considered as natural combinatorial peptide libraries structurally constrained by the cystine-knot scaffold and head-to-tail cyclization.[2,10] Cyclotides can also be chemically synthesized, thus allowing the introduction of specific chemical modifications or biophysical probes.[11–14] More importantly, cyclotides can now be biosynthesized in *E. coli* cells through a biomimetic approach that involves the use of modified protein splicing units[15,16] (Figure 2). This therefore makes them ideal scaffolds for molecular evolution strategies to enable the generation and selection of compounds with optimal binding and inhibitory characteristics against particular molecular targets.

Investigation of the contribution of individual residues to the structural integrity and biological activities of particular cyclotides is crucial for their use in any potential pharmaceutical application.[17] A better understanding of the structural limitations of the cyclotide scaffold, so that sequence modifications in structurally important regions are avoided, can greatly assist in the correct design of cyclotide-based libraries for molecular screening and the selection of de novo sequences with new biological activities or in developing grafted analogues for use as peptide-based drugs[14,18]. Understanding the molecular basis for bioactivity might also allow the minimization or avoidance of undesirable properties, such as cytotoxicity or hemolytic activity, that are found in some cyclotides.[17]

The cyclotides MCoTI-I/II are powerful trypsin inhibitors that have been recently isolated from the dormant seeds of *Momordica cochinchinensis*, a plant member of Cucurbitaceae family.[19] Although MCoTI cyclotides do not share significant sequence homology with other cyclotides beyond the presence of the three cystine bridges, solution NMR has shown that they adopt a similar backbone-cyclic cystine-knot topology[20,21] (Figure 1 A). MCoTI cyclotides, however, share a high sequence homology with related cystine-knot trypsin inhibitors found in squash such as EETI, and it is likely they have a similar binding to that of the EETI-family (Figure 1 B).[19] Hence, cyclic MCoTIs represent interesting candidates for drug design, either through changing their specificity of inhibition or through using their structure as natural scaffolds possessing new binding activities.

Results and Discussion

In the current study we report the biosynthesis and screening of biological activity of libraries based on the cyclotide MCoTI-I. These libraries were designed to contain multiple MCoTI-I mutants, in which all the residues in loops 1–5, except for the Cys residues involved in the cystine-knot, were replaced by different types of amino acid. These mutations included the introduction of neutral (Ala), flexible and small (Gly), hydrophilic (Ser and Thr), hydrophobic (Met and Val), constrained (Pro) and aromatic (Tyr and Trp) residues (see Table 1). The only residue in loop 6 that was mutated was Val1. This residue is a hydrophobic β-branched amino acid highly conserved in other

[a] *W. Wang, Prof. J. A. Camarero*
University of Southern California, Department of Pharmacology and Pharmaceutical Sciences, School of Pharmacy
Los Angeles, CA 90033 (USA)
Fax: (+1) 323-224-7473
E-mail: jcamarer@usc.edu

[b] *J. Austin, Prof. J. A. Camarero*
Lawrence Livermore National Laboratory
Livermore, CA 94550 (USA)

[c] *S. Puttamadappa, Prof. A. Shekhtman*
State University of New York, Department of Chemistry
Albany, NY 12222 (USA)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.200900534.

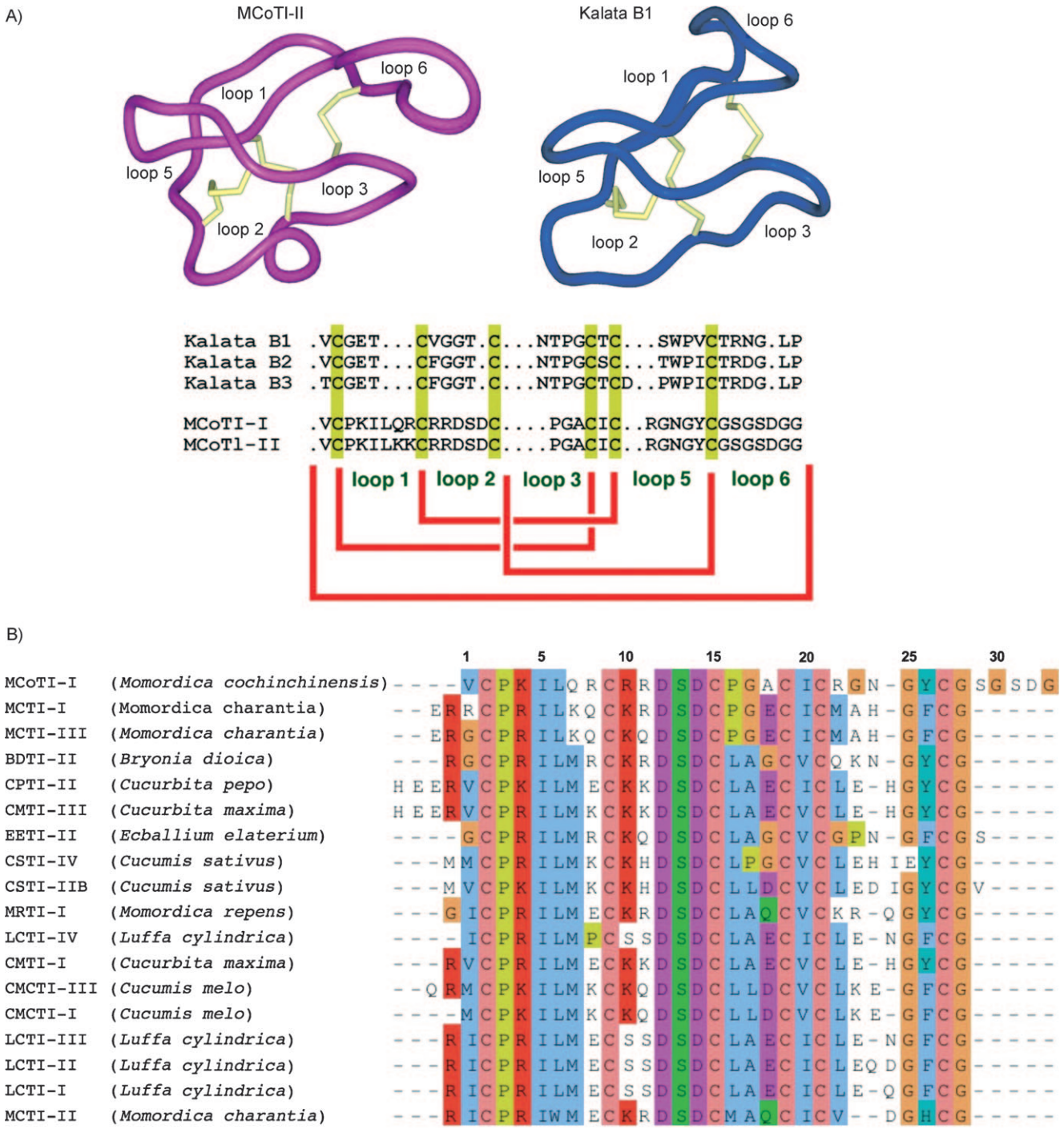


Figure 1. A) Primary and tertiary structures of MCoTI and Kalata cyclotides isolated from *Momordica cochinchinensis* and *Oldenlandia affinis*, respectively.[6,20,21] B) Multiple sequence alignment of cyclotide MCoTI-I with other squash trypsin inhibitors. Multiple sequence alignment was performed by using TCoffee (http://ca.expasy.org/cgi-bin/hub) and visualized by using Jalview.[31]

squash trypsin inhibitors (STIs) and is in close proximity to Lys4 in loop 1, which is responsible for MCoTI's ability to inhibit trypsin. The rest of the residues in loop 6 are not required for folding or biological activity in linear STIs[22] and therefore were not explored. It is believed that loop 6 acts as a very flexible linker to allow cyclization.[23] To our knowledge this is the first time that a cyclotide-based library has been biosynthesized in *E. coli* cells, and a complete amino acid scanning was carried out in the MCoTI-I to explore the effects of individual amino acids on biological activity and structural requirements.

The biosynthesis of MCoTI-I mutants was carried out by using a protein splicing unit in combination with an in-cell intramolecular native chemical ligation reaction (NCL) (Figure 2).[15,16] Intramolecular NCL requires the presence of an N-terminal Cys residue and a C-terminal α-thioester group in the same linear precursor molecule.[24,25] For this purpose, the

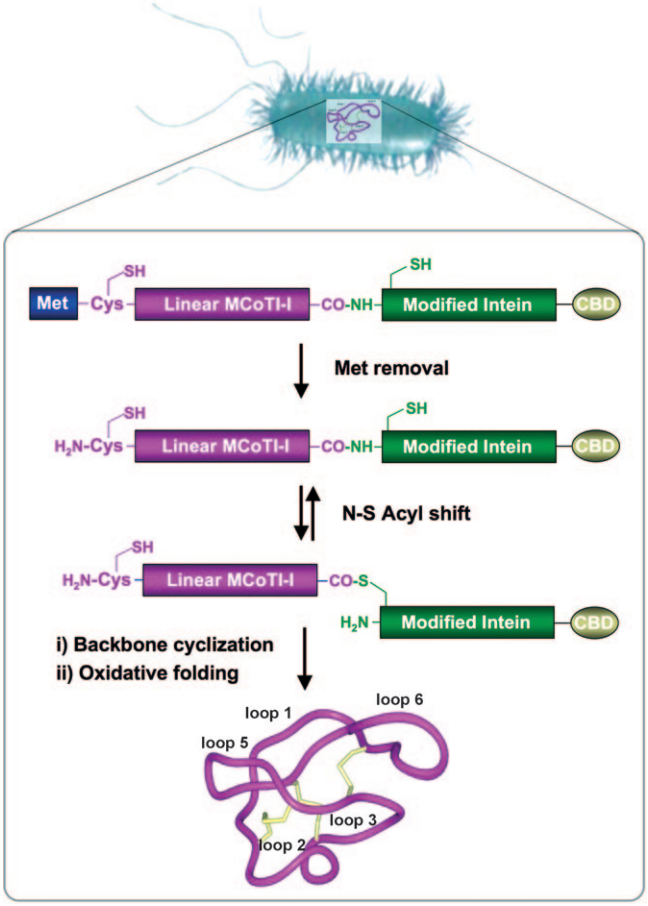


Figure 2. Biosynthetic approach to the production of cyclotide MCoTI-I libraries inside living *E. coli* cells. Backbone cyclization of the linear precursor is mediated by a modified protein splicing unit or intein. The cyclized peptide then folds spontaneously in the bacterial cytoplasm.

MCoTI-I linear precursors were fused in frame at their C and N termini to a modified Mxe gyrase A intein and a Met residue, respectively. This allows the generation of the required C-terminal thioester and N-terminal Cys residue after in vivo processing by endogenous Met aminopeptidase (MAP). We used the native Cys located at the beginning of loop 6 to facilitate the cyclization. This linear construct has been shown to give very good expression and cyclization yields in vivo.[16]

In order to facilitate the analysis and processing of all the mutants, two libraries (Lib1 and Lib2) were produced that contained 13 and 15 different MCoTI-I mutants, respectively (see Table 1). These libraries were designed to contain mutants that could be easily identified by ES-MS. In both libraries, the MCoTI-I wild-type (wt) sequence was included as a control. Synthetic dsDNA fragments encoding the different MCoTI-I mutants were ligated into plasmid pTXB1 in frame with Mxe gyrase intein (Table S1 in the Supporting Information). The resulting plasmid libraries were transformed into competent DH5α *E. coli* cells to give approximately 10^4 colonies (data not shown). All the colonies were pooled, and the corresponding plasmid library was transformed into *E. coli* Origami2(DE3) for protein overexpression.

Expression of the library in *E. coli* produced the corresponding MCoTI mutant–gyrase intein linear fusion precursors with similar yields to that of the wild-type MCoTI-I.[16] The level of in vivo cleavage was estimated to be $\approx$80% following induction for 20 h at 20 °C (Figure S1). These expression conditions maximize the in vivo processing of the linear intein fusion precursors to give natively folded MCoTI cyclotides.[16] In vivo cleavage and processing of the corresponding intein linear fusion precursor can be reduced by induction at slightly elevated temperatures for short times (e.g., 30 °C for 2–4 h) while keeping a similar level of protein expression (Figure S1). This allowed us to vary the amounts of folded MCoTI mutants produced to access different screening methods. For in vitro screening, cyclization can be accomplished in vitro under controlled conditions, and therefore short induction times at relatively higher temperatures will yield more uncleaved linear precursor. Alternatively, in vivo cyclization yields can be easily maximized by using longer induction times and lower induction temperatures (e.g., 20 °C for 20 h) for high throughput in vivo screening.

In order to characterize the MCoTI-based libraries and assess the structural integrity of the MCoTI mutants, we used the biological activity of MCoTI to bind trypsin. Purified MCoTI mutant–gyrase fusion proteins were obtained from *E. coli* Origami2(DE3) cells that were induced at 30 °C for 4 h. Under these conditions, only $\approx$30% of the intein linear precursors were processed in vivo. The fusion precursors were cleaved and cyclized in phosphate buffer (pH 7.2) containing 50 mM glutathione (GSH) for 36 h. In our hands, GSH has been shown to be more effective than other thiols in promoting cyclization and native-like folding of cyclotides and other disulfide-containing peptides in vitro.[15,16,26] This treatment resulted in nearly 100% cleavage of the intein precursors. The soluble fractions were purified on trypsin–sepharose beads, and the bound fractions were analyzed by HPLC and ES-MS to determine the relative presence of the library members able to bind trypsin (Figure 3 A).

As anticipated, the MCoTI-K4A mutant was not found in the trypsin-bound fraction. This residue determines binding affinity and specificity, and can only be replaced by Arg to maintain biological activity.[19] Analysis of the cyclization reaction before affinity purification confirmed the presence of this mutant in the corresponding library (Figure S2). The K4A mutant was also individually cyclized, purified and characterized by NMR, and showed a native cyclotide topology when to compared to MCoTI-I wt (Figure S3 and Table S2), therefore indicating that the lack of biological activity of this mutant was due to the replacement of Lys4 by Ala, and not to the adoption of a non-native fold.

The mutant MCoTI-I G25P was also absent in the trypsin-bound fraction. In vitro cyclization of this mutant revealed that the intein precursor of this cyclotide was not processed efficiently, and the resulting cyclotide was not able to fold properly. Only traces of natively folded G25P were detected in the GSH-induced cyclization/folding of the corresponding intein precursor (Figure S4). The inefficient cleavage of this mutant precursor could be explained by the proximity of a Pro residue

Table 1. Sequences and molecular weights found for the different MCoTI-I mutants used in this work.

Name	Sequence	Molecular weight [Da] Expected	Found
Lib1			
wt	CGSGSDGGVCPKILQRCRRDSDCPGACICRGNGY	3480.94	3480.4 ± 0.6
P3A	CGSGSDGGVC**A**KILQRCRRDSDCPGACICRGNGY	3454.91	3454.2 ± 0.3
K4A	CGSGSDGGVCP**A**ILQRCRRDSDCPGACICRGNGY	3423.85	3423.7 ± 0.6
I5T	CGSGSDGGVCPK**T**LQRCRRDSDCPGACICRGNGY	3468.89	3468.0 ± 0.1
L6A	CGSGSDGGVCPKI**A**QRCRRDSDCPGACICRGNGY	3438.86	3438.7 ± 1.0
Q7G	CGSGSDGGVCPKIL**G**RCRRDSDCPGACICRGNGY	3409.87	3408.5 ± 1.2
Q7M	CGSGSDGGVCPKIL**M**RCRRDSDCPGACICRGNGY	3484.01	3483.7 ± 0.6
R8A	CGSGSDGGVCPKILQ**A**CRRDSDCPGACICRGNGY	3395.84	3396.0 ± 0.1
D12A	CGSGSDGGVCPKILQRCRR**A**SDCPGACICRGNGY	3436.93	3435.7 ± 0.6
S13A	CGSGSDGGVCPKILQRCRRD**A**DCPGACICRGNGY	3464.95	3465.0 ± 1.0
G17A	CGSGSDGGVCPKILQRCRRDSDCP**A**ACICRGNGY	3494.97	3495.0 ± 1.0
A18G	CGSGSDGGVCPKILQRCRRDSDCPG**G**CICRGNGY	3466.92	3467.0 ± 1.7
Y26A	CGSGSDGGVCPKILQRCRRDSDCPGACICRGNG**A**	3388.85	3388.3 ± 1.5
Lib2			
wt	CGSGSDGGVCPKILQRCRRDSDCPGACICRGNGY	3480.94	3480.4 ± 0.6
V1A	CGSGSDGG**A**CPKILQRCRRDSDCPGACICRGNGY	3452.89	3453.0 ± 1.0
V1S	CGSGSDGG**S**CPKILQRCRRDSDCPGACICRGNGY	3468.89	3468.7 ± 1.2
R10A	CGSGSDGGVCPKILQRC**A**RDSDCPGACICRGNGY	3395.84	3396.0 ± 0.1
R10G	CGSGSDGGVCPKILQRC**G**RDSDCPGACICRGNGY	3381.81	3380.3 ± 0.6
R11V	CGSGSDGGVCPKILQRCR**V**DSDCPGACICRGNGY	3423.89	3423.7 ± 0.6
D14A	CGSGSDGGVCPKILQRCRRDS**A**CPGACICRGNGY	3436.93	3436.7 ± 1.2
P16A	CGSGSDGGVCPKILQRCRRDSDC**A**GACICRGNGY	3454.91	3455.0 ± 1.7
I20G	CGSGSDGGVCPKILQRCRRDSDCPGAC**G**CRGNGY	3424.84	3423.7 ± 0.6
R22W	CGSGSDGGVCPKILQRCRRDSDCPGACIC**W**GNGY	3510.97	3510.7 ± 1.2
G23A	CGSGSDGGVCPKILQRCRRDSDCPGACICR**A**NGY	3494.97	3495.0 ± 1.0
N24W	CGSGSDGGVCPKILQRCRRDSDCPGACICRG**W**GY	3553.05	3552.7 ± 1.2
G25P	CGSGSDGGVCPKILQRCRRDSDCPGACICRGN**P**Y	3521.01	3521.0 ± 1.4
Y26A	CGSGSDGGVCPKILQRCRRDSDCPGACICRGNG**A**	3388.44	3388.3 ± 1.5
Y26W	CGSGSDGGVCPKILQRCRRDSDCPGACICRGNG**W**	3503.98	3503.3 ± 1.2

to the MCoTI–intein junction, which might affect the ability of the gyrase intein to produce the thioester intermediate required for the intramolecular cyclization. The Gly25 residue is located on loop 5 and is extremely well conserved in the STIs (Figure 1B), thus corroborating the importance of this residue for correct folding of MCoTI cyclotides.

Remarkably, the remaining mutants were identified on the trypsin-bound fraction, thus indicating that they were able to adopt a native-like structure whose ability to bind trypsin was not significantly disrupted. All the active mutants besides I20G were produced with similar yields to the MCoTI wt (within 50% of the average value), as quantified by HPLC and ES-MS. The folded I20G mutant abundance was estimated to be ≈10% of the average. Cleavage and cyclization of I20G with GSH revealed that, although the thiol-induced cleavage was very efficient, the correctly folded mutant was produced in very low yield (Figure S5); this indicated the importance of this residue for efficient folding in MCoTI cyclotides. In fact, this residue, which is located between the Cys residues at the end and beginning of loops 3 and 5, respectively, is well conserved among the different linear STIs and cyclotides; this shows a preference for β-branched residues and hydrophilic residues at this position. Interestingly, the correctly folded I20G mutant was able to bind trypsin beads, thus confirming the mutant's ability to adopt a native folded structure.

Next, we screened the biological activity of the MCoTI-I libraries produced in vivo. For this purpose both libraries (Lib1 and Lib2) were expressed in *E. coli* Origami2(DE3) cells at 20 °C for 20 h in order to maximize the intracellular processing and folding of the different intein precursors. After the cells had been lysed by sonication, the cellular supernatant was purified by using trypsin-sepharose under competing binding conditions, as described above. The different fractions were then analyzed and quantified by HPLC and ES-MS. The results obtained were very similar to those found with in vitro cyclized libraries (data not shown), thus indicating that the composition of the libraries obtained in vitro and in vivo were practically identical.

In order to establish the relative affinities of the different mutants that are able to bind trypsin versus MCoTI-I wild type, in vitro and in vivo cyclized libraries were incubated with trypsin–sepharose under competing conditions, that is, using only ≈20% of the required trypsin–sepharose beads for stoichiometric binding. This process ensured that cyclotides with tighter affinities competed for binding to trypsin leaving the members of the library with weaker affinities in the supernatant (i.e., unbound fraction). This supernatant was then purified again by using the same approach to extract the remaining active cyclotides. This process was repeated several times until all the active cyclotides found in a particular library sample were completed extracted. This process ensured that cyclotides with tighter affinities for trypsin were extracted during the first affinity purifications leaving the library members with weaker affinities to be purified later on in this sequential extraction process. All the different trypsin-bound fractions were then analyzed and quantified by HPLC and ES-MS (Figure 3B). The results, summarized in Figure 4, show that with MCoTI-I wt as internal reference, the mutants N24W and R22W were consistently able to compete slightly with the rest of mutants (including wt), thus indicating a somewhat tighter affinity for trypsin than the wt sequence. Most of the remaining mutants: P3A, Q7M, R8A, R10A, R10G, R11V, D12A, S13A, D14A, P16A, G17A, A18G, G23A, Y26A and Y26W showed similar elution profiles to that of wt indicating a similar affinity for trypsin. Mutants V1A, V1S, I5T, L6A and Q7G, on the other hand, were consistently extracted after MCoTI wt; this indicates a lower affinity for trypsin than the wt sequence. I20G (not shown) was

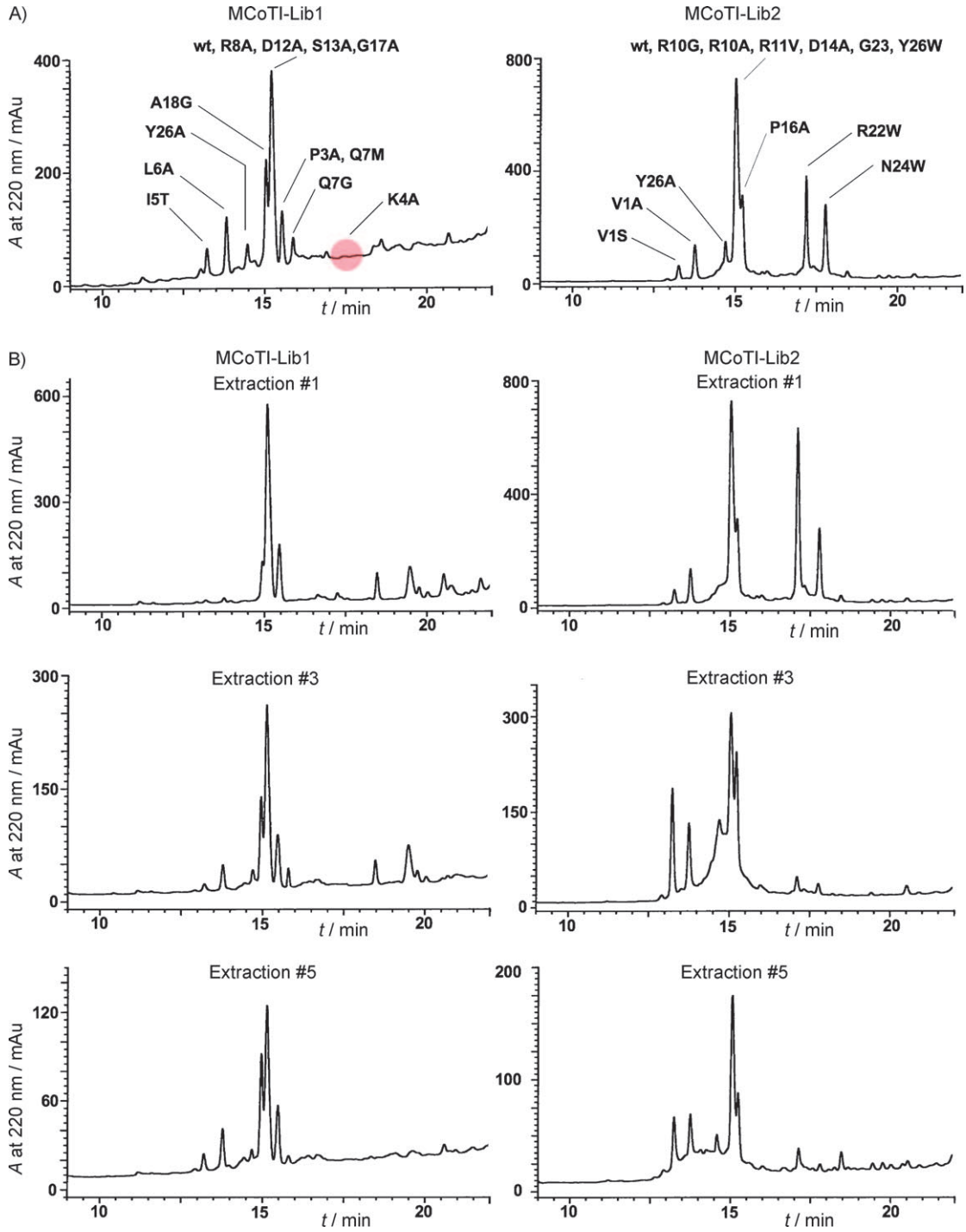


Figure 3. Analytical reversed-phase HPLC traces of trypsin-bound fractions from MCoTI Lib1 (left) and Lib2 (right) libraries obtained in vitro by GSH-induced cleavage and folding. A) Total trypsin-bound fractions. The position where mutant K4A should be eluting is shown in red. B) Sequential fractions extracted during competitive trypsin-binding experiments (see text for detailed description).

also extracted after MCoTI-I wt; however, this could be due to the low abundance of folded cyclotide.

Although there is no structure available for the complex between MCoTI cyclotides and trypsin, the structure of several complexes formed between different STIs and trypsin have been reported so far.[27,28] Based on the high sequence homology between these trypsin inhibitors and MCoTI cyclotides (Figure 1), it is reasonable to assume that they possess the same binding mode to trypsin.[19] Therefore, it is not surprising that mutant K4A was not able to bind trypsin, since K4 is critical for binding to the trypsin specificity pocket.[19] Other mutations in loop 1 also negatively affected trypsin binding. Hence, mutants I5T, L6A and Q7G were consistently eluted in the later fractions in our competing binding experiments, thus indicating a weaker affinity for trypsin. The sequence Lys/Arg-Ile-Leu in loop 1 is extremely well conserved in all linear STIs; this sug-

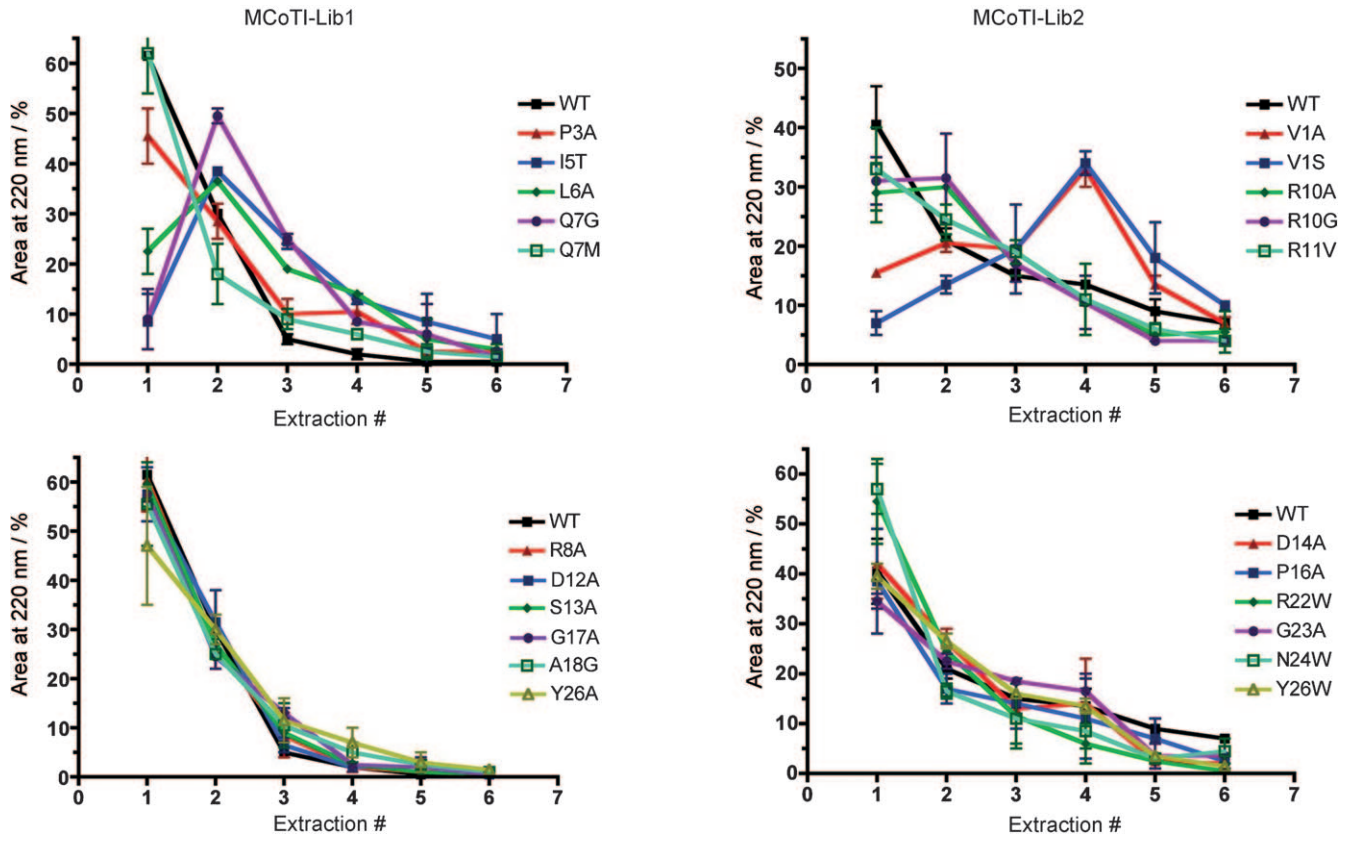


Figure 4. Elution profiles for members of the MCoTI Lib1 (left) and Lib2 (right) extracted by using trypsin-sepharose beads under competing conditions. The results shown are the average data obtained in vivo and in vitro (vertical bars indicate standard deviation). Quantification was done by integration of the corresponding HPLC peaks monitored at 220 nm (Figure 3B). ES-MS was used to calculate the ratio of different cyclotides present in peaks with multiple products or not well resolved by HPLC. The area estimated for mutants with loss or gain of aromatic residues was corrected accordingly to take this into account.[32]

gests that it is required for efficient trypsin binding. Position 7, on the other hand, seems to be more promiscuous and it is able to accept hydrophobic residues (Q7M showed a similar elution pattern to MCoTI-wt) and positively charged residues (cyclotide MCoTI-II has a Lys residue in this position), but not a small and flexible residue like Gly (mutant Q7G shows weaker trypsin affinity than wt). Also in loop 1, mutation R8A did not significantly affect trypsin binding, and the corresponding mutant showed an elution pattern similar to that of wt. In agreement with this result, this position is not especially well conserved in linear STIs allowing the presence of charged (both positively and negatively) and Pro residues.

All the mutations explored in loop 2 had similar elution patterns to the wt sequence. This should be expected since this loop is solvent exposed and on the opposite side to loop 1. The only mutation affecting trypsin binding in loop 3 was represented by mutant P16A. The rest of the mutants in this loop behaved similarly to the wt sequence. This loop is partially exposed in the structure of several linear STIs with trypsin and it shows significant sequence heterogeneity among the different STIs. Position 16, however, is usually occupied in other STIs by hydrophobic residues (mainly Leu and Met); this could explain the observed behavior of mutant P16A.

None of the mutations in loop 5, besides G25P, had an adverse effect on trypsin binding. It is interesting to remark that mutants Y26A and Y26W showed a similar elution profile to the wt sequence (Figure 4). This position is very well conserved among different STIs, being occupied mainly by either aromatic (Tyr or Phe) or in some cases Ile and His. Analysis of the structure of linear *Cucurbita pepo* trypsin inhibitor-II (CPTI-II, which shares ≈75% sequence homology with MCoTI-I) complexed with bovine trypsin[27] shows that this position makes a direct contact with the trypsin Tyr151 residue, which is highly conserved among different trypsin homologues. Intriguingly, mutants R22W and N24W seemed to slightly outcompete the rest of the library members including MCoTI-I wt in our binding competing experiments (Figures 3B and 4). These residues are in close proximity to Tyr26 and they could help to further stabilize the aromatic interaction described before between the MCoTI-I mutants and trypsin.

The position corresponding to Val1 at the end of loop 6 was also explored by including mutants V1A and V1S in this study. This position favors the presence of hydrophobic residues (mainly Val, Met and Ile) among different STIs, although hydrophilic and charged residues are also found in some STIs. Visual inspection of the CPTI-II–trypsin complex,[27] for example, reveals that this position is in close proximity to the trypsin Trp215. This aromatic residue is highly conserved among the different trypsin homologues, thus indicating that this interaction might be important to the stabilization of the complex.

Consistent with this finding, replacement of Val1 in MCoTI-I by Ser and Ala produced mutants that consistently showed a weaker affinity than MCoTI-I wt (Figure 5).

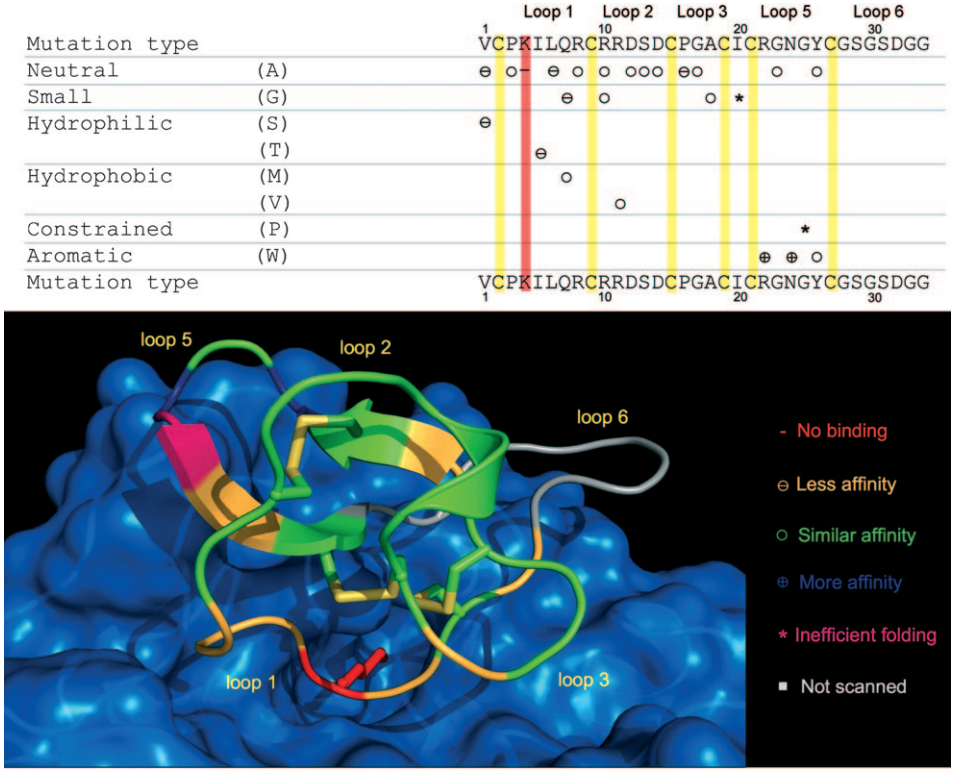


Figure 5. Summary of the relative affinities for trypsin of the different MCoTI-I mutants studied in this work. A model of cyclotide MCoTI-I bound to trypsin is shown at the bottom and indicates the positions of the mutations. The Lys4 side chain is shown in red bound to the specificity pocket of trypsin. The model was produced by homology modeling at the Swiss model workspace (http://swissmodel.expasy.org//SWISS-MODEL.html)[33] by using the structure of CPTI-II-trypsin complex (PDB ID: 2btc)[27] as the template. The structure was generated by using the PyMol software package.

In summary, these data provide significant insights into the structural constraints of the MCoTI cyclotide framework and the functional elements for trypsin binding. To our knowledge, this is the first time that the biosynthesis of a genetically encoded library of MCoTI-based cyclotides containing a comprehensive suite of amino acid mutants has been reported. Craik and co-workers have also recently reported the chemical synthesis of a complete suite of Ala mutants for the cyclotide Kalata B1 (KB-1).[17] These mutants were fully characterized structurally and functionally. Their results indicated that only two of the mutations explored (KB-1 W20A and P21A, both located in loop 5, see Figure 1) prevented folding.[17] The mutagenesis results obtained in our work show similar results, thus highlighting the extreme robustness of the cyclotide scaffold to mutations. Only two of the 27 mutations studied in cyclotide MCoTI-I, G25P and I20G, negatively affected the adoption of a native cyclotide fold. Intriguingly, the rest of the mutations allowed the adoption of a native fold, as indicated by ES-MS analysis and their ability to bind trypsin (or NMR in the case of K4A). These results should provide an excellent starting point for the effective design of MCoTI-based cyclotide libraries for rapid screening and selection of de novo cyclotide sequences with specific biological activities.

The libraries used in this work were produced either in vitro by GSH-induced cyclization/folding or by in vivo self-processing of the corresponding precursor proteins. In both cases the results were similar; this indicates that this approach is quite general for the production of complex libraries. Importantly, the in vivo biosynthesis of cyclotide-based libraries could have tremendous potential for drug discovery. This study shows that MCoTI cyclotides could provide an ideal scaffold for the biosynthesis of large combinatorial libraries inside living *E. coli* cells. Coupled to an appropriate in vivo reporter system, this library could rapidly be screened by using high-throughput technologies such as fluorescence-activated cell sorting.[29,30]

See the Supporting Information for experimental details.

Acknowledgements

This work was supported by funding from the School of Pharmacy at the University of Southern California and Lawrence Livermore National Laboratory.

Keywords: cyclotides · expressed protein ligation · genetically encoded libraries · inhibitors · protein splicing

[1] D. J. Craik, S. Simonsen, N. L. Daly, *Curr. Opin. Drug Discovery Dev.* **2002**, *5*, 251.
[2] D. J. Craik, M. Cemazar, C. K. Wang, N. L. Daly, *Biopolymers* **2006**, *84*, 250.
[3] R. J. Clark, N. L. Daly, D. J. Craik, *Biochem. J.* **2006**, *394*, 85.
[4] D. J. Craik, M. Cemazar, N. L. Daly, *Curr. Opin. Drug Discovery Dev.* **2006**, *9*, 251.
[5] M. L. Colgrave, D. J. Craik, *Biochemistry* **2004**, *43*, 5965.
[6] O. Saether, D. J. Craik, I. D. Campbell, K. Sletten, J. Juul, D. G. Norman, *Biochemistry* **1995**, *34*, 4147.
[7] K. P. Greenwood, N. L. Daly, D. L. Brown, J. L. Stow, D. J. Craik, *Int. J. Biochem. Cell Biol.* **2007**, *39*, 2252.
[8] D. J. Craik, R. J. Clark, N. L. Daly, *Expert Opin. Invest. Drugs* **2007**, *16*, 595.
[9] H. Kolmar, *Curr. Opin. Pharmacol.* **2009**; DOI: 10.1016/j.coph.2009.05.004.
[10] D. C. Ireland, M. L. Colgrave, N. L. Daly, D. J. Craik, *Adv. Exp. Med. Biol.* **2009**, *611*, 477.
[11] N. L. Daly, S. Love, P. F. Alewood, D. J. Craik, *Biochemistry* **1999**, *38*, 10606.
[12] O. Avrutina, H. U. Schmoldt, H. Kolmar, U. Diederichsen, *Eur. J. Org. Chem.* **2004**, 4931.

[13] P. Thongyoo, E. W. Tate, R. J. Leatherbarrow, *Chem. Commun.* **2006**, 2848.
[14] P. Thongyoo, N. Roque-Rosell, R. J. Leatherbarrow, E. W. Tate, *Org. Biomol. Chem.* **2008**, *6*, 1462.
[15] R. H. Kimura, A. T. Tran, J. A. Camarero, *Angew. Chem.* **2006**, *118*, 987; *Angew. Chem. Int. Ed.* **2006**, *45*, 973.
[16] J. A. Camarero, R. H. Kimura, Y. H. Woo, A. Shekhtman, J. Cantor, *ChemBioChem* **2007**, *8*, 1363.
[17] S. M. Simonsen, L. Sando, K. J. Rosengren, C. K. Wang, M. L. Colgrave, N. L. Daly, D. J. Craik, *J. Biol. Chem.* **2008**, *283*, 9805.
[18] D. J. Craik, N. L. Daly, J. Mulvenna, M. R. Plan, M. Trabi, *Curr. Protein Pept. Sci.* **2004**, *5*, 297.
[19] J.-F. Hernandez, J. Gagnon, L. Chiche, T. M. Nguyen, J.-P. Andrieu, A. Heitz, T. T. Hong, T. T. C. Pham, D. Le Nguyen, *Biochemistry* **2000**, *39*, 5722.
[20] A. Heitz, J.-F. Hernandez, J. Gagnon, T. T. Hong, T. T. C. Pham, T. M. Nguyen, D. Le-Nguyen, L. Chiche, *Biochemistry* **2001**, *40*, 7973.
[21] M. E. Felizmenio-Quimio, N. L. Daly, D. J. Craik, *J. Biol. Chem.* **2001**, *276*, 22875.
[22] O. Avrutina, H. U. Schmoldt, D. Gabrijelcic-Geiger, D. Le Nguyen, C. P. Sommerhoff, U. Diederichsen, H. Kolmar, *Biol. Chem.* **2005**, *386*, 1301.
[23] A. Heitz, O. Avrutina, D. Le-Nguyen, U. Diederichsen, J.-F. Hernandez, J. Gracy, H. Kolmar, L. Chiche, *BMC Struct. Biol.* **2008**, *8*, 54.
[24] a) J. A. Camarero, J. Pavel, T. W. Muir, *Angew. Chem.* **1998**, *110*, 361; *Angew. Chem. Int. Ed. Engl.* **1998**, *37*, 347.
[25] J. A. Camarero, T. W. Muir, *J. Am. Chem. Soc.* **1999**, *121*, 5597.
[26] J. Austin, R. H. Kimura, Y. H. Woo, J. A. Camarero, *Amino Acids* **2009**; DOI: 10.1007s00726-009-0338-0334.
[27] R. Helland, G. I. Berglund, J. Otlewski, W. Apostoluk, O. A. Andersen, N. P. Willassen, A. O. Smalas, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1999**, *55*, 139.
[28] R. Krätzner, J. É. Debreczeni, T. Pape, T. R. Schneider, A. Wentzel, H. Kolmar, G. M. Sheldrick, I. Uson, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2005**, *61*, 1255.
[29] R. H. Kimura, E. R. Steenblock, J. A. Camarero, *Anal. Biochem.* **2007**, *369*, 60.
[30] H. Sancheti, J. A. Camarero, *Adv. Drug Delivery Rev.* **2009**; DOI: 10.1016/j.addr.2009.07.003.
[31] M. Clamp, J. Cuff, S. M. Searle, G. J. Barton, *Bioinformatics* **2004**, *20*, 426.
[32] B. J. Kuipers, H. Gruppen, *J. Agric. Food Chem.* **2007**, *55*, 5445.
[33] N. Guex, M. C. Peitsch, *Electrophoresis* **1997**, *18*, 2714.

Received: August 25, 2009
Published online on September 24, 2009

Supplemental Information

General materials and methods

Analytical HPLC was performed on a HP1100 series instrument with 220 nm and 280 nm detection using a Vydac C18 column (5 μm, 4.6 x 150 mm) at a flow rate of 1 mL/min. Semipreparative HPLC was performed on a Waters Delta Prep system fitted with a Waters 2487 Ultraviolet-Visible (UV-vis) detector using a Vydac C18 column (15-20 μm, 10 x 250 mm) at a flow rate of 5 mL/min. All runs used linear gradients of 0.1% aqueous trifluoroacetic acid (TFA, solvent A) vs. 0.1% TFA, 90% acetonitrile in H_2O (solvent B). UV-vis spectroscopy was carried out on an Agilent 8453 diode array spectrophotometer, and fluorescence analysis on a Jobin Yvon Flurolog-3 spectroflurometer. Electrospray mass spectrometry (ES-MS) analysis was routinely applied to all cyclized peptides. ES-MS was performed on a Sciex API-150EX single quadrupole electrospray mass spectrometer, MS/MS was performed on an Applied Biosystems API 3000 triple quadrupole mass spectrometer. Calculated masses were obtained by using ProMac v1.5.3. Protein samples were analyzed by SDS-PAGE. Samples were run on Invitrogen (Carlsbad, CA) 4-20% Tris-Glycine Gels. The gels were then stained with Pierce (Rockford, IL) Gelcode Blue, photographed/digitized using a Kodak (Rochester, NY) EDAS 290, and quantified using NIH Image-J software (http://rsb.info.nih.gov/ij/). DNA sequencing was performed by Davis Sequencing (Davis, CA) or DNA Sequencing and Genetic Analysis Core Facility at the Unviersity of Southern California using an ABI 3730 DNA sequencer, and the sequence data was analyzed with DNAStar (Madison, WI) Lasergene v5.5.2. All chemicals were obtained from Sigma-Aldrich (Milwaukee, WI) unless otherwise indicated.

Construction of expression plasmids

Plasmids expressing the MCoTI-I precursors were constructed using the pTXB1 expression plasmids (New England Biolabs), which contain an engineered Mxe Gyrase intein, respectively, and a chitin-binding domain (CBD). Oligonucleotides coding for the MCoTI-I wild type and mutant sequences (Table S1) were synthesized, phosphorylated and PAGE purified by IDT DNA (Coralville, IA). Complementary strands were annealed in 0.3 M NaCl and the resulting double stranded DNA (dsDNA) was purified using Qiagen's (Valencia, CA) miniprep column and buffer PN. pTXB1 plasmids was double digested with NdeI and SapI (NEB). The linearized vectors and the MCoTI-I encoding dsDNA fragments were ligated at 15°C overnight using T4 DNA Ligase (New England Biolabs). The ligated plasmids were transformed into DH5α cells (Invitrogen) and plated on Luria Broth (LB)-agar containing ampicillin. Positive colonies were grown in 5 mL LB containing ampicillin at 37°C overnight and the corresponding plasmids purified using a Miniprep Kit (Qiagen). Plasmids were initially screened by EcoRI digestion, as this restriction site is removed during cloning. Preliminary positives were expressed (see below) and fully characterized by ES-MS.

Expression and purification of recombinant proteins

Origami(DE3) or Origami2(DE3) cells (Novagen, San Diego, CA) were transformed with the MCoTI-I plasmids (see above). Expression was carried out in LB medium (1-2 L) containing ampicillin at room temperature or 30°C for 2 h or overnight (20 h), respectively. Briefly, 5 mL of an overnight starter culture derived from either a single clone or single plate (Ala-scan library) were used to inoculate 1 L of LB media. Cells were grown to an OD at 600 nm of ≈0.5 at 37°C, and expression was induced by the addition of isopropyl-β -D-thiogalactopyranoside (IPTG) to a final concentration of 0.3 mM at the temperatures and times mentioned above and in the manuscript. The cells

were then harvested by centrifugation. For fusion protein purification, the cells were resuspended in 30 mL of lysis buffer (0.1 mM EDTA, 1 mM PMSF, 50 mM sodium phosphate, 250 mM NaCl buffer at pH 7.2 containing 5% glycerol) and lysed by sonication. The lysate was clarified by centrifugation at 15,000 rpm in a Sorval SS-34 rotor for 30 min. The clarified supernatant was incubated with chitin-beads (2 mL beads/L cells, New England Biolabs), previously equilibrated with column buffer (0.1 mM EDTA, 50 mM sodium phosphate, 250 mM NaCl buffer at pH 7.2) at 4°C for 1 h with gentle rocking. The beads were extensively washed with 50 bead-volumes of column buffer containing 0.1% Triton X100 and then rinsed and equilibrated with 50 bead-volumes of column buffer. *In-vivo* cleavage was quantified by SDS-PAGE analysis of the purified fusion proteins using the NIH Image-J software package.

Concomitant cleavage, cyclization and folding of MCoTI-I cyclotides with GSH. Purified MCoTI–Intein-CBD fusion proteins were cleaved with 50 mM GSH in degassed column buffer. The cyclization/folding reactions were kept for up to 2 days at 25°C with gentle rocking. For small scale reactions, aliquots was taken each day (when necessary) and analyzed by HPLC. The reduced and oxidized circular MCoTI-I cyclotides were analyzed by ES-MS (Table 2). The supernatant of the cyclization reaction was separated by filtration and the beads were washed with additional column buffer (1 column volume per each mL of beads). The supernatant and washes were pooled, and the oxidized-circular peptides were typically purified by semipreparative HPLC using a linear gradient of 15-45% solvent B over 30 min.

Purification of MCoTI-I based libraries using trypsin-sepharose beads. Preparation of trypsin-sepharose beads: NHS-activated Sepharose was washed with 15 volumes of ice-cold 1 mM HCl. Each volume of beads was incubated with an equal volume of coupling buffer (50 mM NaCl, 200 mM sodium phosphate buffer at pH 6.0) containing 2 mg of Porcine Pancreas Trypsin type IX-S (14,000 units/mg) for 3 h with

gentle rocking at room temperature. The beads were then rinsed with 10 volumes of coupling buffer, and incubated with excess coupling buffer containing 100 mM ethanolamine (Eastman Kodak) for 3 h with gentle rocking at room temperature. Finally, the beads were washed with 50 volumes of wash buffer (200 mM sodium acetate buffer at pH 3, 250 mM NaCl) and stored in one volume of wash buffer.

≈30 mL of clarified lysates (in vivo obtained libraries) or 10 mL of GSH-induced cyclization/folding reaction mixture (in vitro obtained libraries) were typically incubated with 1.0 mL of trypsin-sepharose for one hour at room temperature with gentle rocking, and centrifuged at 3000 rpm for 1 min. The beads were washed with 50 volumes of PBS containing 0.1% Triton X100, then rinsed with 50 volumes of PBS, and drained of excess PBS. Bound peptides were eluted with 2.0 mL of 8 M GdmHCl and fractions were analyzed by RP-HPLC and ES-MS/MS.

Competing trypsin-binding experiments

Competing binding experiments were performed as described above but using 0.2 mL of trypsin-sepharose beads instead. For every library sample, up to 6 sequential extractions were performed. All competing trypsin-binding experiments were performed by duplicate.

Recombinant expression of ^{15}N-labeled MCoTI-I wt and K4A

^{15}N-labeled cyclotides were produced by GSH-induced cleavage of the intein precursors in vitro as described above. Expression of intein precursors was accomplished as described above but growing the cells in M9 minimal medium containing 0.1% $^{15}NH_4Cl$. Folded cyclotides were purified by semipreparative HPLC using a linear gradient of 15-45% solvent B over 30 min. The isolated yield for both purified MCoTI cyclotides was around 0.3 mg/L. Purified products were characterized by HPLC and ES-MS (Fig. S5) and 2D-NMR (Figure S2 and Table S2). Total yield was ≈0.5 mg of folded cyclotide per liter of bacterial culture.

NMR characterization of MCoTI-I wt and K4A mutant

NMR samples were prepared by dissolving either [*U*-, ^{15}N] MCoTI-I or [*U*-, ^{15}N] K4A MCoTI-I into 90% H_2O/10% ^2H_2O (v/v) or 100% D_2O to a concentration of approximately 0.2 mM with the pH adjusted to 3.4 by addition of dilute HCl. All ^{1}H NMR data were recorded on Bruker Avance II 700 MHz spectrometer equipped with a cryoprobe. Data were acquired at 27 °C, and 2,2-dimethyl-2-silapentane-5-sulfonate, DSS, was used as an internal reference. All 3D experiments, ^{1}H{^{15}N}-TOCSY-HSQC and ^{1}H{^{15}N}-NOESY, were performed according to standard procedures[1] with spectral widths of 12 ppm in proton dimensions and 35 ppm in nitrogen dimension. The carrier frequency was centered on the water signal, and the solvent was suppressed by using WATERGATE pulse sequence. TOCSY (spin lock time 80 ms) and NOESY (mixing time 150 ms) spectra were collected using 1024 t_3 points, 256 t_2 anf 128 t_1 blocks of 16 transients. Spectra were processed using Topspin 1.3 (Bruker). Each 3D-data set was apodized by 90^0-shifted sinebell-squared in all dimensions, and zero filled to 1024 x 512 x 256 points prior to Fourier transformation. Assigments for the backbone nitrogens, H^{α} and H' protons (Figure S2 and Table S2 and S3) of folded MCoTI-I wt and mutant K4A were obtained using standard procedures.[1, 2]

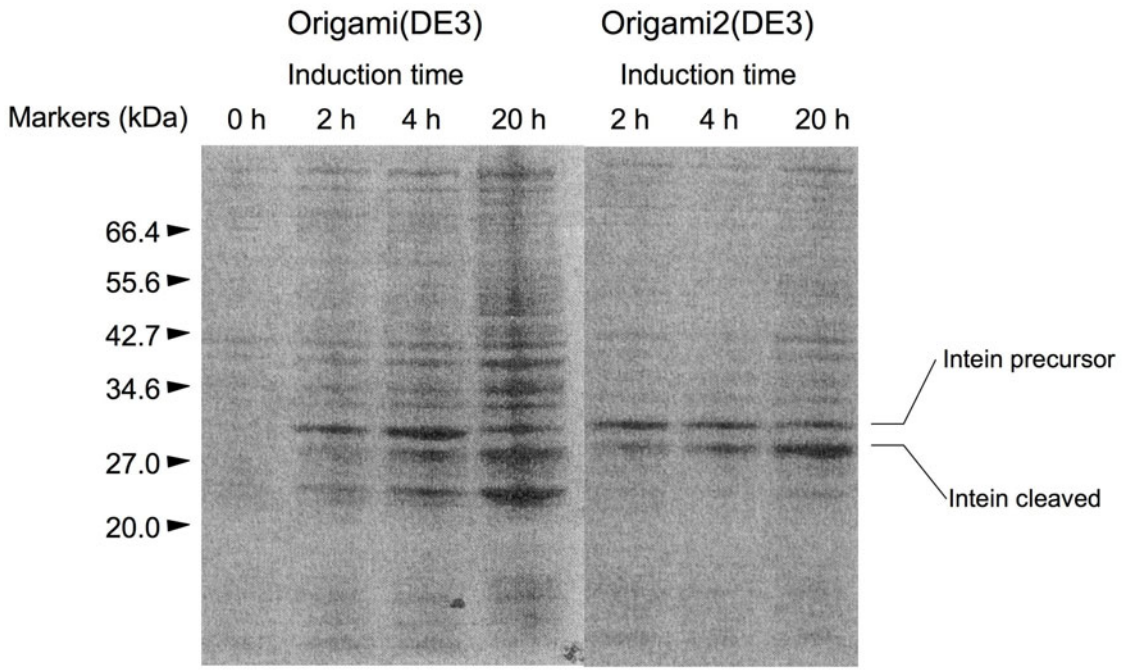


Figure S0. Analysis by 4-20% gradient SDS-PAGE of the expression levels and in vivo cleavage of MCoTI-Lib2 precursors using different cellular backgrounds. Expression levels for MCoTI-Lib1 were similar. Induction was carried out at 30°C (2 and 4 h) or 20°C (20 h) by adding 0.3 mM IPTG.

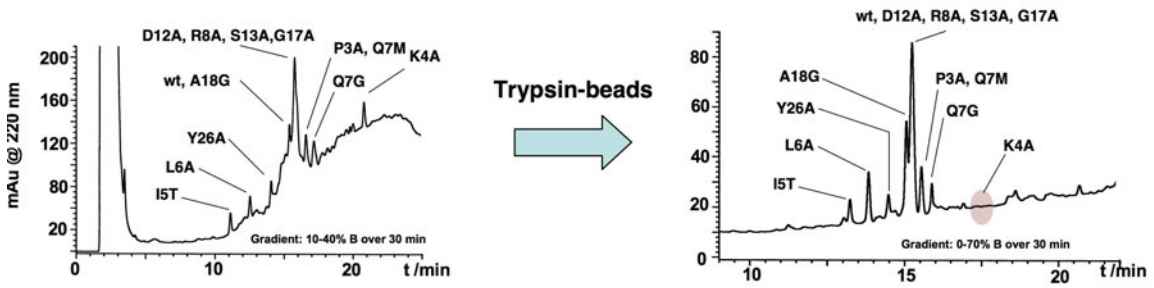


Figure S1. Analytical reversed-phase HPLC trace of GSH-induced cyclization of MCoTI-Lib1 precursors before and after being purified by affinity chromatography on trypsin-sepharose. Identification of the different MCoTI-I mutants was carried out by ES-MS.

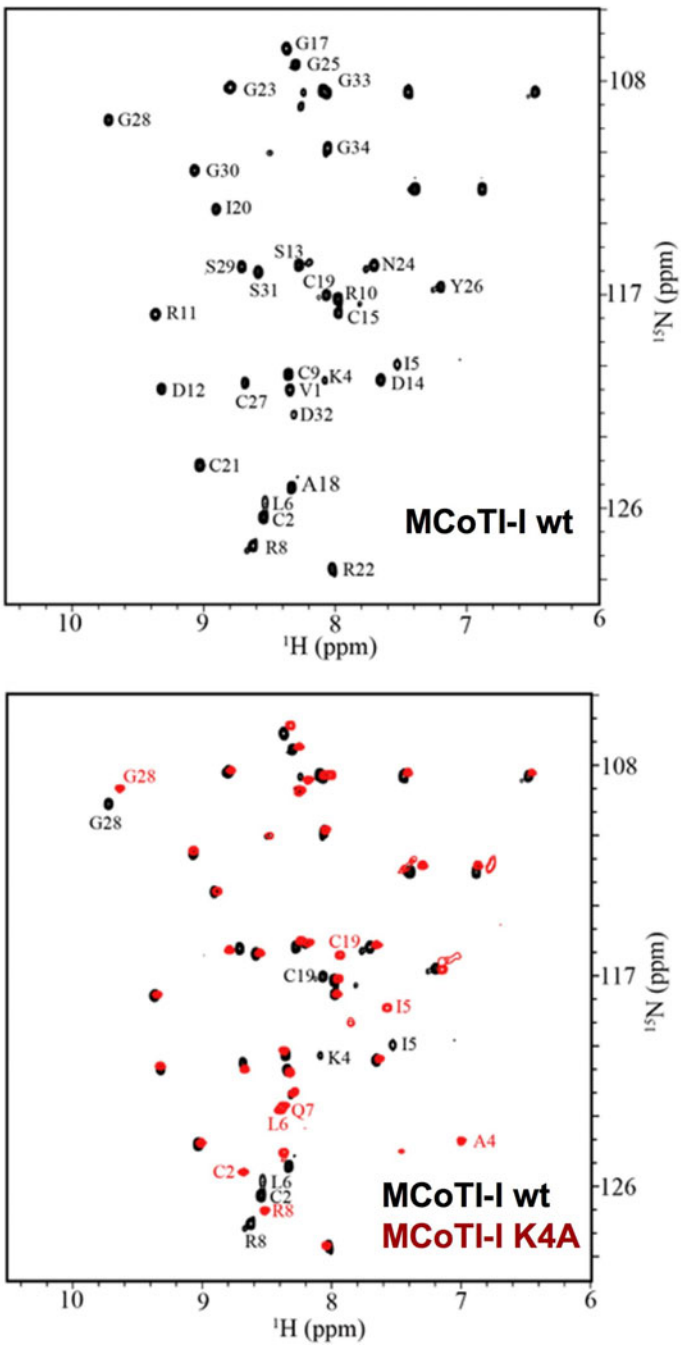


Figure S2. Heteronuclear ^{1}H{^{15}N} HSQC-spectra of recombinant wt MCoTI-I wt (black) and K4A mutant (red). Chemical shift assignments of the wt MCoTI-I amino acid residues are shown in black. K4A mutant residues that changed chemical shifts by more that 0.3 ppm in proton dimension are labeled in red. Small unassigned peaks in both wt and K4A spectra of MCoTI-I are from a minor isomer of the protein due to a known isomerization of the backbone at an Asp-Gly sequence in loop 6 of MCoTI-I. NMR experiments were acquired on a Bruker Avance II 700 MHz NMR spectrometer equipped with a cryoprove at 27^{o} C. NMR samples of 0.2 mM of [*U*-, ^{15}N] MCoTI-I and 0.25 mM of

[*U*-, ^{15}N] K4A MCoTI-I were in 90% H_2O/10% D_2O adjusted to pH 3.5 by addition of dilute HCl.

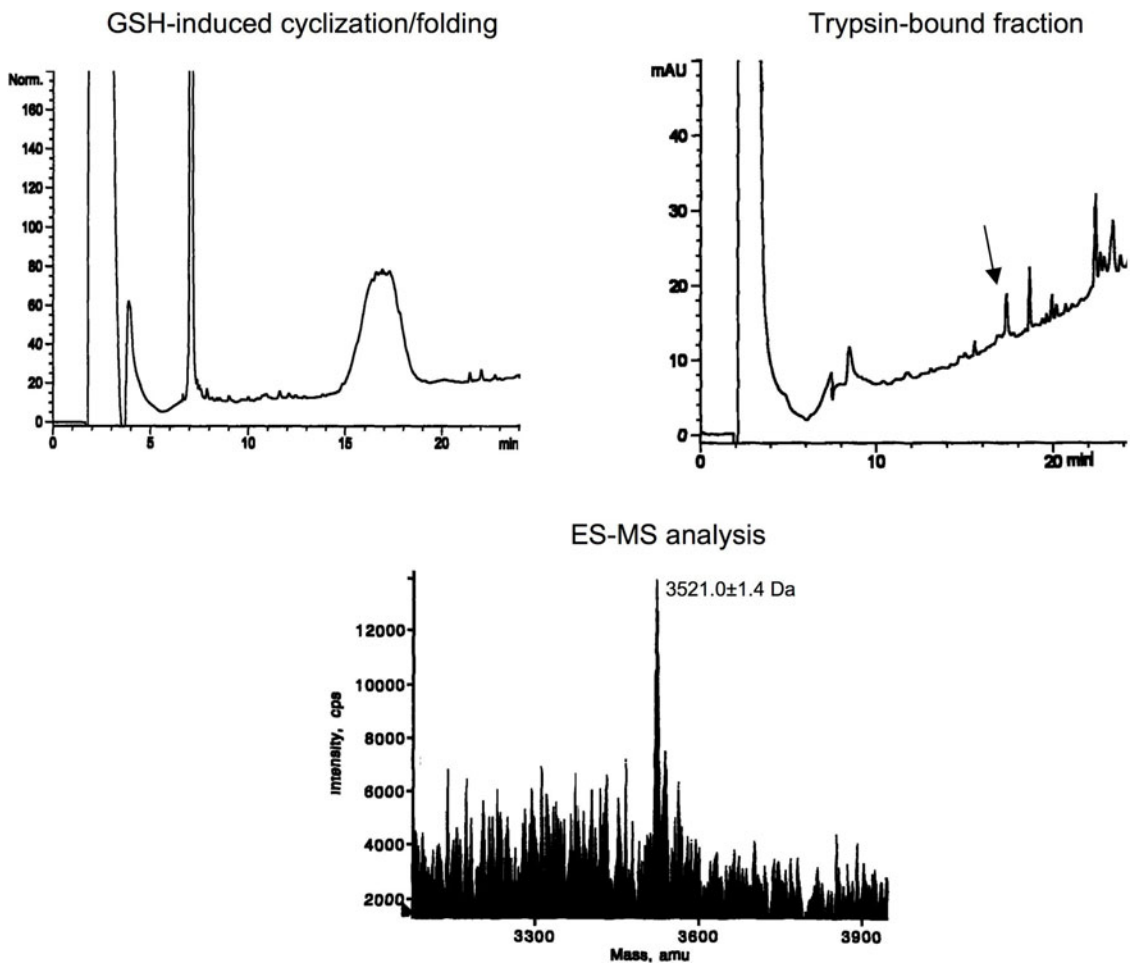


Figure S3. Analytical reversed-phase HPLC trace of GSH-induced cyclization of MCoTI-G25P precursor before and after being purified by affinity chromatography on trypsin-sepharose. Identification of the folded mutant was carried out by ES-MS.

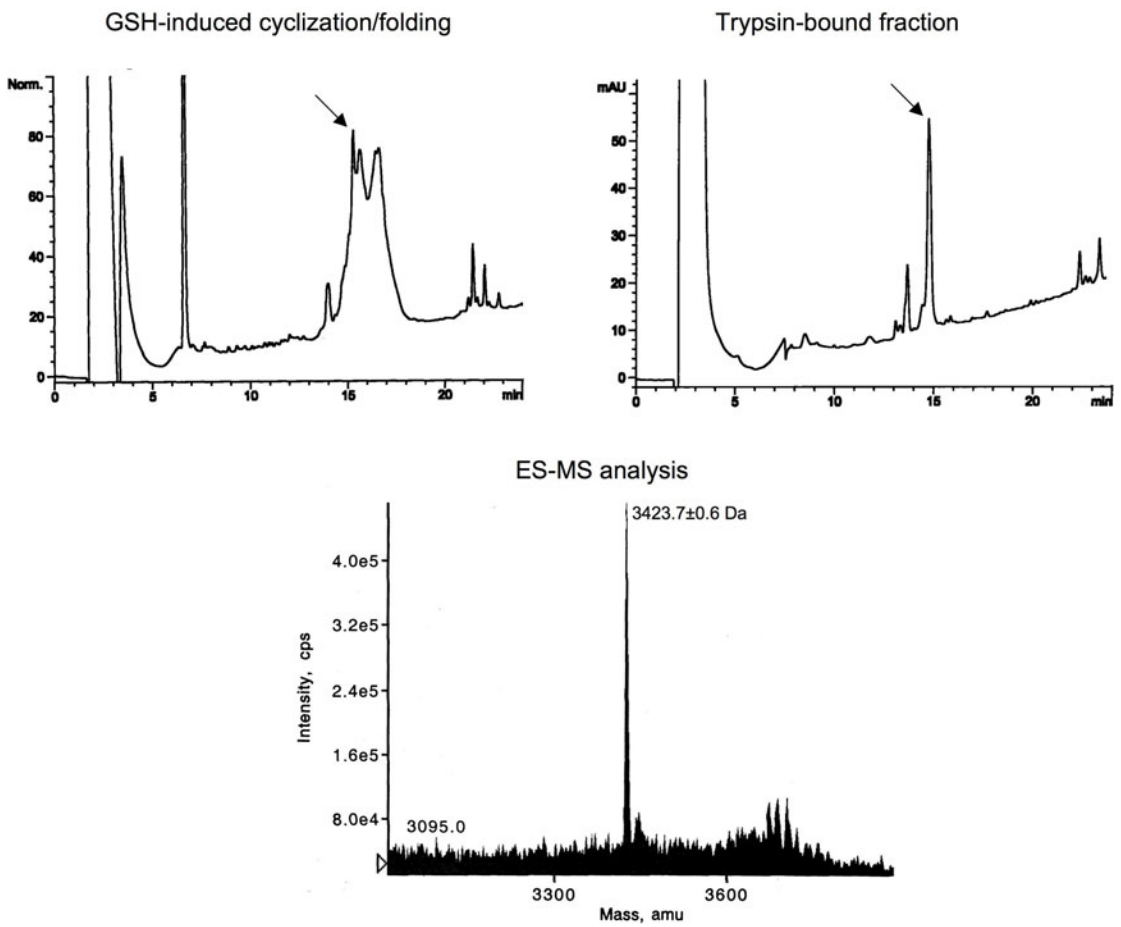


Figure S4. Analytical reversed-phase HPLC trace of GSH-induced cyclization of MCoTI-I20G precursor before and after being purified by affinity chromatography on trypsin-sepharose. Identification of the folded mutant was carried out by ES-MS.

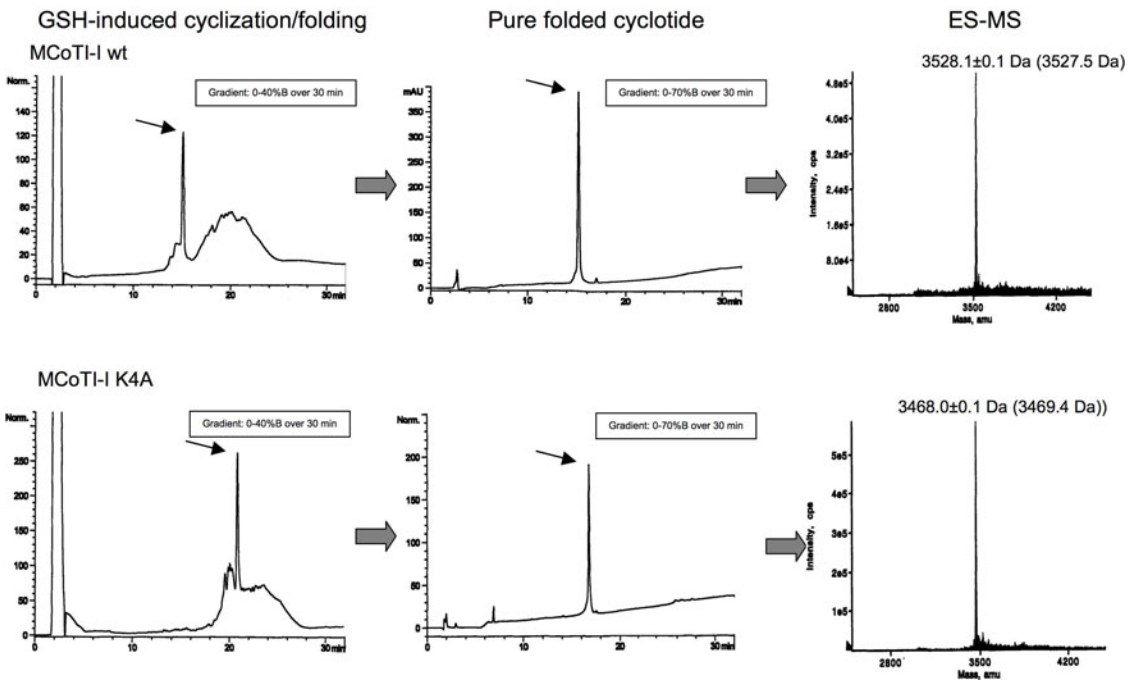


Figure S5. Analytical reversed-phase HPLC trace of GSH-induced cyclization of ^{15}N labeled MCoTI-I wt and K4A precursors before and after being purified by preparative HPLC. Identification of the folded mutant was carried out by ES-MS. Expected molecular weight for the ^{15}N-labeled cyclotide is shown in parenthesis.

Table S1. Forward (p5) and reverse (p3) 5'-phosphorylated oligonucleotides used to clone the different MCoTI-1-intein linear precursors into the pTXB1 expression plasmid.

Cyclotide name		Oligonucleotide sequence
wt	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcc gtcgtgactctgactgcccgggtgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcacccgggcagtcagagtcacgacggc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
V1A	p5	5′-TATGtgcggttctggttctgacggtggtgcttgcccgaaaatcctgcagcgttgcc gtcgtgactctgactgcccgggtgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcacccgggcagtcagagtcacgacggc aacgctgcaggattttcgggcaagcaccaccgtcagaaccagaaccgcaCA-3′
V1S	p5	5′-TATGtgcggttctggttctgacggtggttcttgcccgaaaatcctgcagcgttgcc gtcgtgactctgactgcccgggtgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcacccgggcagtcagagtcacgacggc aacgctgcaggattttcgggcaagaaccaccgtcagaaccagaaccgcaCA-3′
P3A	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcgctaaaatcctgcagcgttgcc gtcgtgactctgactgcccgggtgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcacccgggcagtcagagtcacgacggc aacgctgcaggattttagcgcaaacaccaccgtcagaaccagaaccgcaCA-3′
K4A	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccggctatcctgcagcgttgcc gtcgtgactctgactgcccgggtgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcacccgggcagtcagagtcacgacggc aacgctgcaggatagccgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
I5T	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaaccctgcagcgttgcc gtcgtgactctgactgcccgggtgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcacccgggcagtcagagtcacgacggc aacgctgcagggttttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
L6A	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcgctcagcgttgcc gtcgtgactctgactgcccgggtgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcacccgggcagtcagagtcacgacggc aacgctgagcgattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
Q7G	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgggtcgttgcc gtcgtgactctgactgcccgggtgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcacccgggcagtcagagtcacgacggc aacgacccaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
Q7M	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgatgcgttgcc gtcgtgactctgactgcccgggtgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcacccgggcagtcagagtcacgacggc aacgcatcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
R8A	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcaggcttgcc gtcgtgactctgactgcccgggtgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcacccgggcagtcagagtcacgacggc aagcctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
R10A	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcg ctcgtgactctgactgcccgggtgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcacccgggcagtcagagtcacgagcgc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
R10G	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcg gtcgtgactctgactgcccgggtgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcacccgggcagtcagagtcacgaccgc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
R11V	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcc gtgttgactctgactgcccgggtgcttgcatctgccgtggtaacggttac -3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcacccgggcagtcagagtcaacacggc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA -3′

D12A	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcc gtcgtgcttctgactgcccgggtgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcacccgggcagtcagaagcacgacggc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
S13A	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcc gtcgtgacgctgactgcccgggtgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcacccgggcagtcagcgtcacgacggc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
D14A	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcc gtcgtgactctgcttgcccgggtgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcacccgggcaagcagagtcacgacggc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
P16A	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcc gtcgtgactctgactgcgctggtgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcaccagcgcagtcagagtcacgacggc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
G17A	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcc gtcgtgactctgactgcccggctgcttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaagcagccgggcagtcagagtcacgacggc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
A18G	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcc gtcgtgactctgactgcccgggtggttgcatctgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcagatgcaaccacccgggcagtcagagtcacgacggc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
I20G	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcc gtcgtgactctgactgcccgggtgcttgcggttgccgtggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccacggcaaccgcaagcacccgggcagtcagagtcacgacggc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
R22W	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcc gtcgtgactctgactgcccgggtgcttgcatctgctggggtaacggttac-3′
	p3	5′-GCAgtaaccgttaccccagcagatgcaagcacccgggcagtcagagtcacgacggc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
G23A	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcc gtcgtgactctgactgcccgggtgcttgcatctgccgtgctaacggttac-3′
	p3	5′-GCAgtaaccgttagcacggcagatgcaagcacccgggcagtcagagtcacgacggc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
N24W	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcc gtcgtgactctgactgcccgggtgcttgcatctgccgtggttggggttac-3′
	p3	5′-GCAgtaaccccaaccacggcagatgcaagcacccgggcagtcagagtcacgacggc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
G25P	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcc gtcgtgactctgactgcccgggtgcttgcatctgccgtggtaacccgtac-3′
	p3	5′-GCAgtacgggttaccacggcagatgcaagcacccgggcagtcagagtcacgacggc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
Y26W	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcc gtcgtgactctgactgcccgggtgcttgcatctgccgtggtaacggttgg-3′
	p3	5′-GCAccaaccgttaccacggcagatgcaagcacccgggcagtcagagtcacgacggc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′
Y26A	p5	5′-TATGtgcggttctggttctgacggtggtgtttgcccgaaaatcctgcagcgttgcc gtcgtgactctgactgcccgggtgcttgcatctgccgtggtaacggtgct-3′
	p3	5′-GCAagcaccgttaccacggcagatgcaagcacccgggcagtcagagtcacgacggc aacgctgcaggattttcgggcaaacaccaccgtcagaaccagaaccgcaCA-3′

Table S2. Summary of the ^{1}H and $^{15}N^{\alpha}$ NMR assignments for the main-chain protons (i.e. N^{α}-H and C^{α}-H) of recombinant MCoTI-I.

Residues	^{1}H-$^{15}N^{\alpha}$ (ppm)	$^{15}N^{\alpha}$ (ppm)	^{1}H-C^{α} (ppm)
Val 1	8.33	120.96	3.89
Cys 2	8.54	126.39	4.92
Pro 3			
Lys 4	8.10	120.61	4.17
Ile 5	7.523	119.968	4.25
Lue 6	8.53	125.78	4.35
Gln 7			
Arg 8	8.61	127.62	4.37
Cys 9	8.34	120.34	4.49
Arg 10	7.97	117.20	4.58
Arg 11	9.36	117.83	4.35
Asp 12	9.32	120.98	4.76
Ser 13	8.27	115.75	4.68
Asp 14	7.65	120.57	4.44
Cys 15	7.97	117.74	4.82
Pro 16			
Gly 17	8.37	106.63	3.65
Ala 18	8.33	125.14	4.31
Cys 19	8.06	117.00	4.50
Ile 20	8.9	113.42	4.24
Cys 21	9.03	124.20	3.95
Arg 22	8.02	128.58	4.17
Gly 23	8.79	108.27	3.80
Asn 24	7.70	115.77	4.52
Gly 25	8.30	107.30	3.82
Tyr 26	7.19	116.66	5.10
Cys 27	8.68	120.72	5.19
Gly 28	9.72	109.62	4.36, 3.73
Ser 29	8.71	115.78	4.34
Gly 30	9.06	117.73	3.71, 4.22
Ser 31	8.58	116.05	4.27
Asp 32	8.31	122.07	4.49
Gly 33	8.08	10844	3.67, 3.89
Gly 34	8.05	110.83	3.67, 4.14

Table S3. Summary of the ^{1}H and $^{15}N^{\alpha}$ NMR assignments for the main-chain protons (i.e. N^{α}-H and C^{α}-H) of recombinant MCoTI-I K4A.

Residues	^{1}H-$^{15}N^{\alpha}$ (ppm)	^{15}N (ppm)	^{1}H-C^{α} (ppm)
Val 1	8.32	121.17	3.80
Cys 2	8.68	125.40	4.86
Pro 3			
Ala 4	6.70	124.03	4.44
Ile 5	7.57	118.41	4.21
Lue 6	8.40	122.79	4.27
Gln 7	8.36	122.55	4.30
Arg 8	8.51	127.02	4.30
Cys 9	8.36	120.17	
Arg 10	7.95	117.17	4.58
Arg 11	9.34	117.80	4.27
Asp 12	9.32	120.91	4.81
Ser 13	8.23	115.57	4.67
Asp 14	7.63	120.54	4.40
Cys 15	7.96	117.81	4.77
Pro 16			
Gly 17	8.31	106.30	3.65
Ala 18	8.36	124.53	4.31
Cys 19	7.93	116.09	4.63
Ile 20	8.88	113.42	4.19
Cys 21	9.006	124.15	3.93
Arg 22	8.04	128.53	4.17
Gly 23	8.77	108.24	3.80
Asn 24	7.65	115.71	4.64
Gly 25	8.25	107.23	3.55, 3.89
Tyr 26	7.14	116.71	5.10
Cys 27	8.67	121.05	5.13
Gly 28	9.63	108.99	3.75, 4.30
Ser 29	8.78	115.94	4.37
Gly 30	9.06	111.70	3.72, 4.22
Ser 31	8.55	116.07	4.28
Asp 32	8.29	122.04	4.49
Gly 33	8.01	108.48	3.67
Gly 34	8.04	110.76	4.65

References

[1] J. Cavanagh, M. Rance, *J. Magn. Res.* **1992**, *96*, 670.
[2] K. Wuthrich, *NMR of Proteins and Nucleic Acids.*, **1986**.